SMALL AREA PROJECTIONS: MODELLING THE DEMOGRAPHIC AND EPIDEMIOLOGICAL DYNAMICS OF A RURAL AREA IN SOUTH AFRICA

By
Makandwe Nyirenda

A dissertation submitted to the faculty of Commerce of the University of Cape Town in partial fulfilment of the requirements for the degree of master of philosophy in demography

Centre for Actuarial Research
December 2006

DEDICATION

This report is dedicated to all my friends who made me feel at home in a 'strange' land.

ACKNOWLEDGEMENTS

I would like to acknowledge the tireless efforts of my supervisor, Dr T. Moultrie, for his invaluable contribution to this study. I would like to thank him for the patience and understanding shown towards me. I would also like to thank him for the extra hours put in to ensure this project was completed.

I also would like to acknowledge the invaluable contribution and advice received from Professor Rob Dorrington, even when called upon at short notice.

Recognition should also be made of the Africa Centre for Health and Population Studies for allowing me access to their datasets. I also pay tribute to all the people at the Africa Centre who in one way or another contributed to this project. In a very special way I also thank the people of Hlabisa district who are so ever willing to provide data to the Africa Centre.

Lastly, but as the saying goes, certainly not the least, I wish to thank the Andrew Mellon Foundation and the University of Cape Town, Postgraduate Funding Office for making a financial contribution to my studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The project's objective is to derive a robust base population for use in the ASSA model to replicate the Africa Centre Demographic Information System (ACDIS) data. The problem at hand is to identify a robust means to derive a base population given the inadequacies of the available potential data sources. The major sources of data are the 2001 census and the ACDIS. The former data source is however shown not to be reliable for setting the base population of the study area, while the latter source does not go far back in time to cover the chosen base year. The study uses a deterministic model that incorporates cohort-component methods. This model is arguably more robust than available alternative models. The study demonstrates that the ACDIS population can be reasonably replicated on certain key variables. On other variables, however, the model does not perform well. The major limitation of the study was availability and quality of data. In general, despite a number of limitations, the model produces results that show reasonable consistency with the observed population.

1 INTRODUCTION

1.1 Background

This desire to know what the future is likely to be is partly the reason population projections or forecasts continue to be demanded of demographers. When individuals and institutions plan for the future they usually want to base their decisions on some highly probable future outlook. As Keyfitz notes "knowledge is better than ignorance" (Keyfitz, 1982: 729). It is therefore important to constantly test and develop more robust methods for making approximations of the future.

This study explores the application of the widely used cohort-components population projection methods (Booth, 2006) and a projection model that combines demographic and epidemiological projections (Johnson and Dorrington, 2006) in modelling the population dynamics of a South African sub-district.

This study will engage with the practical and methodological issues of small-area population projections. It seeks to establish as robust as possible a base population for population projection of an area in Hlabisa district, Kwa-Zulu Natal province. The established base population will then be imported into the ASSA demographic and epidemiological projection model. The aim of the study is to use the model to replicate the population dynamics of the Africa Centre Demographic Information System (ACDIS) as closely as possible. The techniques to be adopted here are seen as highly useful in small-area population projections especially in data-poor settings.

1.2 Objectives of the study

The study has both general and specific objectives. The general objective of this study is to engage with the methodological and practical issues of small-area population projections by applying standard demographic projection techniques to a well-defined population in a well-defined area, using the ASSA demographic and epidemiological model. Specifically the study objectives are first to assess the reliability and accuracy of census data in relation to DSS data as a source of input data for the study areas' population and epidemiological projections; second, to derive as robust as possible an estimate of the population for the ACDIS coverage area by age and sex at base year, 1985, appropriate for use with the ASSA models; and lastly, to calibrate the ASSA model to replicate the observed population in the ACDIS, and in so doing to develop a model that can be used for small-area projections.

1.3 Statement of the problem

In this proposed study a base population has to be developed to model future population trends for a district facing a severe HIV/AIDS pandemic. The ASSA model with which the projections will be carried out requires that the base population be set prior to the start of the epidemic. The model requires an AIDS-free population at the base year, to which it then factors in HIV prevalence and mortality. The problem at hand is that formal data collection at the Africa Centre DSS started only in 2000, (Ngom, Solarsh, Benzler *et al.*, 2002). By this time the AIDS situation had already reached epidemic proportions (Karim and Karim, 2002). This therefore makes the data that has been collected in the ACDIS inappropriate as base population inputs into the model, as it does not go far back enough to cover the base year. In addition the observed population is of a population already severely affected by the HIV pandemic.

In many cases, especially in data-rich populations, projections are carried out using census data to set the base population (Campbell, 1996; Siegel and Swanson, 2004; Booth, 2006). The quality of demographic data in general, and census data in particular, in developing countries has been well documented (Cleland, 1996). Reliable census data for the study area, which is a poor and predominantly rural area, is therefore likely not to exist. Given the South African apartheid government's history of using all institutions they could manage to exert control over the country (Hindson, 1987; Posel, 1991; Muthien, 1994) the 1996 and 2001 censuses may be considered as probably the more reliable source of data for the district under study. Unfortunately, in addition to the typical errors that such data tend to contain, the structure and composition of these data have also been significantly affected by the AIDS epidemic.

The problem at hand is therefore to identify a robust means to derive a base population given the inadequacies of the available potential data sources. This problem is further compounded by the need to establish this base population for a period assumed not to have been affected by HIV/AIDS. Once established the issue will then be, using this population and a demographic and epidemiological model, to replicate the population dynamics for the area under study.

1.4 Justification and significance for the study

The wide availability of computer based projection models and techniques has made the undertaking of population projections relatively easy. However, as the US Census Bureau notes, the complexity is in coming up with input data that are of adequate quality (US Census Bureau, 2005). Fertility, mortality and migration assumptions also need to be made that are consistent with the most likely future trends.

The first justification of undertaking this study was that to the best of knowledge of the author, there has not been any such study in the district in particular and in South Africa in general. It is therefore believed that the study will contribute to the knowledge base of small-area projections. Personal interest in this branch of demography was yet another reason justifying this study.

Further justification for undertaking this study in the area in question was that it will provide valuable information for service delivery, planning and assessing future health and social needs for the area. For instance the potential cost-benefit analysis of an antiretroviral therapy roll out can be informed by running the model to investigate the potential impact of the roll out on survival post-infection. The model will potentially be a tool against which the reliability and accuracy of the ACDIS data can be made.

Compared to national projections this is a relatively under-researched area, further justifying this study, hence the accuracy and bias of such projections is not fully known (Smith and Shahidullah, 1995). This is despite the improvements over the years that have certainly been achieved concerning projection methodology (Booth, 2006).

For small areas, methodological issues and bias which are not present in national projections abound (Smith and Shahidullah, 1995). For instance in many cases both international and internal migration is assumed to be negligible in national projections. For small areas not only does international migration become important but internal migration becomes even more critical (Keyfitz, 1972), and cannot just be assumed to be negligible. Population projections at national level are generally more robust compared to sub-national population projections. The lower the unit of analysis below the national level, the more sensitive the projections are to assumptions made in the base period (Jannuzzi, 2005). The unique features of the study area provide a good opportunity to engage with the major facets of small-area projections. Results of this study will provide a basis for testing the accuracy and reliability of a model that makes use of cohort-component methods.

Even though the study will not go as far as making the actual projection beyond the present date, because the estimation of fertility, mortality and migration assumptions that would be required for such an undertaking are beyond the scope of this study, it will nonetheless provide a very significant starting point for such an undertaking in future by any interested parties.

1.5 Expected outcomes

This study aims to come up with a robust base population that will be used to project the population for Hlabisa district using the ASSA model. To an extent this study will test the ASSA model for robustness in small-area population projections. This should also be of interest to users, and sceptics alike, of the model.

It is expected that findings of this study will help to expand frontiers of knowledge vis-à-vis the under-researched area of small area population projections. In this respect it is expected that another outcome of this study will be to spark off further research projects in small-area projections.

1.6 Chapter outline

The rest of this report is organised in this manner. The next chapter (2) presents the relevant literature around the study topic. This will be followed by chapter 3, which presents the methods and tools used to answer the research objectives of this study. Chapter 4 will then evaluate census data and demographic surveillance data as potential sources of input data for the model to be fitted for the area under study. The next chapter (5) after this will then present the model calibration process. This chapter will also present and discuss some key results from the fitted model. Finally, chapter 6 will end this report with a recap of the major findings and recommendations of this study.

2 LITERATURE REVIEW

2.1 Introduction

In demography a distinction is usually made between a population projection and a population forecast. The latter is a statement of the predicted best possible outcome of population trends at a future date, whereas the former is a population outcome of a set of assumptions (Keyfitz, 1972; Keyfitz, 1981). In essence therefore, a population forecast may be wrong in that the predicted future population may turn out to be very different from the population observed at that future date, while a projection by definition is never wrong, regardless of how the future pans out, provided arithmetical or other errors are avoided in the computation of the specified algorithms (Ahlburg and Lutz, 1998). In the literature and everyday use, these two terms are sometimes used interchangeably. In fact many demographers argue that the distinction between these two terms is at best superfluous (Stoto, 1983; Ahlburg and Lutz, 1998; Booth, 2006). Dorn (1950) and Keyfitz (1972) draw similar conclusions that from a users' viewpoint as long as published numbers purport to be best possible future populations, users will view them as forecasts irrespective of what demographers choose to call them. Dorn furthermore contends (seemingly from a demographers' viewpoint) that making projections is not "merely doing sums in arithmetic for self entertainment; they must be serious and really believe that the projected population estimates were in fact reasonable forecasts or predictions of the future" (Dorn, 1950: 313). The approach taken in this study is to draw a distinction between population forecasts and projections.

Approximations of future human populations have for many years been routinely carried out by researchers, academics, most statistical agencies and also by the United Nations for most nations in the world. Cox (1952) asks why these estimations are continuously undertaken despite them being repeatedly wide off the mark. His answer is that aside from the usually cited reason of being a tool in planning and decision making:

> forecasts for the future have the same attractions as- and often no greater degree of success than- prognostications about the weather, the economic outlook, or even the possibilities of interplanetary travel (Cox, 1952: 82).

That is, the population projection subject is simply of perennial interest.

According to Tayman (1996), comments about the uncertainty and error inherent in population projections still hold true even at present; and demographers still cannot provide guarantees concerning their projection results. In spite of this, population projections continue to be demanded by both public and private agencies, and increasingly for small areas (Smith and Shahidullah, 1995).

Around the middle of the last century, Siegel (1953) argued that given the present knowledge and techniques, accurate forecasting of small-area populations was not possible; a position he said was generally shared even among prominent and learned demographers. The purpose of this chapter and the study as a whole is to explore whether, using current knowledge and tools, the population of a small geographical area can be more robustly projected.

2.2 Definition of concepts

Before exploring the literature about small-area projections it is necessary to briefly define commonly-used concepts and terminologies. First, one needs to define what is meant by "small-area". There is no gold-standard definition of a small-area. Ghosh and Rao (1994) suggest that a small-area is any geographical area below national level. The term may, therefore, refer to any area from a province, state[1], district, county, ward or census enumeration area. In this study the term small-area is used to refer to a sub-district.

A second set of concepts relates to the labelling of time points. The base year is defined as the earliest year from which a projection model is run. The final year of a projection is referred to as the target year. The term launch year is used to refer to the year from which projections start to be made after the last year for which data are known. The interval between the base year and the launch year is defined as the base period. The interval between the launch year and the target year is referred to as the projection horizon (Smith, Tayman and Swanson, 2002; Rayer, Smith and Tayman, 2005).

2.3 Models for small-area projection

The reason the field of small-area population projections is relatively under-researched can be attributed to the many problems that such projections often suffer from. Smith and Shahidullah (1995) mention three major problems of small-area population projections. The first is the availability and quality of data. According to Smith and Shahidullah, the method used to undertake small-area projections is highly dependent

[1] In countries with a federal system of government e.g USA and Nigeria, a state may be taken as the equivalent of a province.

on the available data. Data for comprehensive population projections of subdistricts are usually unavailable because data tend to be tabulated and sometimes only made available for large geographical areas such as provinces or districts. When data are available for a small area they may be based on sample estimation or some other approximations, and hence may contain unacceptably large errors (Ghosh and Rao, 1994). A second major problem of small-area projections is that geographical boundaries are less clear and more fluid at a sub-national level. Where such changes have occurred, historical data have to be adjusted to render them spatially consistent to current data. The third problem Smith and Shahidullah (1995) identify is the greater volatility in population change for small areas compared to large areas that may result from factors such as commercial or residential development, construction of group housing quarters or the occupation of vacant land.

There are four major models that can be used in population projections for small areas: trend extrapolation, apportionment or ratio models, structural, and cohort-component. Trend extrapolation models are the simplest models and involve extending trends observed in the past or the present into the future (Stoto, 1983; Smith, 2006). This can be achieved by a variety of extrapolation techniques such as geometric progression, exponential progression or linear extrapolation. The attraction of these models is the minimal data that they require. The major weakness of these models results from the implicit assumption that historical trends will continue into the future. Since human population is dynamic, these models are less appropriate for projection intervals greater than ten years, as the study by Smith and Sincich (1990) shows. In a study analysing population forecast errors in relation to the length of the base period, Smith and Sincich compare population forecasts derived using three simple extrapolation techniques: linear extrapolation, exponential extrapolation and a shift technique. They find that while linear extrapolation and shift-share techniques produce no significant differences in mean absolute percentage errors (MAPEs)[2] for base periods of a decade or more, for exponential extrapolation techniques there is a significant effect on the MAPEs for forecasts periods longer than 10 years. They conclude that with the simple extrapolation techniques highest possible forecast accuracy is attained with

[2] Forecast error (F_t) was calculated as $F_t=[(\hat{P}_t - P_t)/ P_t]*100$. Where, $\hat{P}_t$ is the forecast population; P_t is the observed population and t is the target year. MAPE is a measure of forecast accuracy and is the average percentage forecast error when sign is ignored (Smith and Sincich, 1990).

approximately 10-year forecast horizons; over the long-run large errors should be expected to be made.

A second class of models that may be used in small-area population projections are apportionment or ratio models (Simpson, Diamond, Tonkin *et al.*, 1996; Siegel and Swanson, 2004; Smith, 2006). These models require that an independent projection be made for a parent population such as the district, province or national population. The relationship of the small area population as a proportion of the parent area population is estimated. The small area's population is then obtained by computing its share of the parent area's population projection (Greenberg, 1972). The apportionment can also be done by more statistically-determined relationships such as a regression formula (Ericksen, 1973; Oshungage, 1986). The independently projected population can be obtained by any one of the four projection models presented here. The major limitation of this method is the assumption that the absolute change in a small area's share of the parent population will remain stable throughout the projection horizon (Smith and Sincich, 1990). In practice, however, the impact of changes in the components of population change is usually bigger for small areas than larger areas. As a result, population growth rates are more volatile the smaller the area (Smith and Shahidullah, 1995). Changes which would normally have little impact on the population structure and distribution of a larger area such as a province or national level, usually have a huge destabilising effect on the population distribution of a smaller area (Jannuzzi, 2005). Over a relatively long projection interval, therefore, the relationship between the parent area and the small area population distribution cannot be expected to remain constant. These models are also more appropriate only for population projections by age and sex. The number of variables or characteristics of the parent population that can be projected and then apportioned to the constituent small areas is limited. Another major disadvantage of the ratio method is that projections for constituent small areas must usually be constrained to sum to the parent area's population. The process of establishing relationships between the parent area and the small areas can be challenging when the parent area is composed of many small units.

Schneider (1956) did a comparison of the ratio method to the cohort-component method. He concludes that though the ratio method is attractive for quick results its greatest disadvantage is the sensitivity of the estimates for the small area to the assumptions made in extrapolating the ratios of the small area to the parent percentage distribution. Jannuzzi (2005) argues that most of the proposed small-area population projections models are data intensive. He therefore, proposes an alternative model which combines population projections carried out using cohort-component methods at regional level and a system of differential equations, to apportion the region's total population into municipal bounds (small areas) of Brazil. According to him, this model is useful only when there is available good census and vital registration data. This model is not useful for this study for several reasons. First, as will be shown in Chapter 4 there is a lack of reliable census data for the study area. Second, the model seems more applicable for total population projections. Third, as Jannuzzi himself notes, the model is apparently more reasonable for projections in a 5 to 10 year horizon. Fourth, the system of equations is taken from ecology equations "used to represent population dynamics of competitive species inside a closed habitat, with limited support capacity" (Jannuzzi, 2005: 3). These equations may lead to biased population estimates because the system implies the parent area is closed to migration. That is, only internal migration within the area is allowed for.

A third group of models that may be used for small-area population projections are structural models. In essence structural models are demographic projection models that take into account the relation between components of population change (that is, fertility, mortality and migration) and socio-economic variables. A renowned example of these models is the World3 model of the 1970s which formed the basis for the theory of limits to growth (O'Neill, Balk, Brickman *et al.*, 2001). Structural models are often applied in conjunction with cohort-component models by incorporating into the projections the likely impact of events such as occupation of vacant land, housing development, establishment of employment opportunities in the area, establishment of institutions like prisons or military bases (Smith, 2006). These models are data-intensive, and the required data may not be readily available for such small areas. In addition the future course of socio-economic variables is thought to be much more difficult to predict compared to demographic processes (Keyfitz, 1981). Further, the causal relationship between the components of population change and socio-economic variables is also generally not uni-directional, making it even harder to incorporate the socio-economic variables in the projections. As such these models are infrequently used.

The final model type that may be used in small-area projections are cohort-component models. Since their first use in 1895 by Edwin Cannan (Cannan, 1978 [1895]), cohort-component methods have become the most widely used methods in projections at both the national and sub-national level (Cohen, 1998; US Census Bureau, 2005; Booth, 2006; Smith, 2006). These models project the population by accounting separately for each of the components of population change. In a cohort-component projection, the population is divided into age and sex groups defined as cohorts. A cohort is simply a group of people sharing common characteristics or experience (Smith, Tayman and Swanson, 2002). For instance, an age cohort refers to people born in the same year or interval of years. The method entails following each cohort until the end of its lifetime or the projection horizon (Hollmann, Mulder and Kallan, 2000; O'Neill, Balk, Brickman *et al.*, 2001), whichever occurs first. This follow-up starts with the population at the base year of the projections. For single-age cohorts, this method entails that those aged x at time T will be aged $x+1$ at time $T+1$. For instance, those aged 0 in 2000 will be aged 1 in 2001; those aged 1 in 2000 will be aged 2 in 2001 and likewise for the other ages. The number of deaths for each age cohort is estimated by exposing the cohort to their respective mortality level. The estimated deaths are then subtracted from the population in the cohort to get survivors into the next age group. The number of migrants into the age cohort $x+1$ are then added to the projected number of survivors.

The projected number of people aged 0 is obtained by first multiplying the assumed or observed age-specific fertility rates by the female population in the childbearing interval, 15-49 years, to get the number of births occurring for each year. The estimated total births are then separated into male and female using the assumed sex ratio at birth. The estimated births by respective sex are then allowed to survive forward by multiplying the numbers by their respective survival probability. In-migrants aged 0 are then added to this number of survivors, while out-migrants are subtracted. This process of updating the population in each age-sex cohort is repeated for each year of the projection horizon to arrive at the projected population by age and sex for each year (O'Neill, Balk, Brickman *et al.*, 2001).

A few studies have been conducted to test the reliability and accuracy of cohort-component models. Smith and Tayman (2003) used projections at national level and for counties in Florida state to evaluate the precision and bias of population projections by age by comparing the accuracy of the state and county projections derived using application of a full cohort-component projection method to projections derived from a simpler variant of the cohort-component method developed by Hamilton and Perry (1962). This simpler variant of the cohort-component requires only data from 2 consecutive censuses. The Hamilton-Perry method can be used to obtain population projections for an area *n*-years after the second census, where n is the intercensal interval. The computation involves calculating cohort-survival rates by dividing the population aged *i* at time of the second census by the population aged *i-n* in the first census. The estimated survival rates are then multiplied by the population aged *i* at time of the second census to get projections for the year *t+n*. The assumption implied in this method is that mortality is constant between the two censuses. The population under n years (that is of children) is derived using child-woman ratios or age-specific fertility rates. Thus, this method can be used to obtain projections for an area 5 or 10 years after the second census, for quinquennial or decennial censuses respectively. Smith and Tayman compare the Mean Absolute Percentage Errors (MAPEs) of this simpler variant of the cohort-component method to the full cohort-component method and show first, that patterns of age errors are different between national and sub-national projections. Second, the magnitude of age errors is not the same across all age groups. Third, as the projection horizon lengthens the margin in age errors narrow.

Application of the cohort-component method requires the population age-sex structure for the base year and estimates for the components of population change (fertility, mortality and migration) for the base period to be made. Estimation of the fertility, mortality and migration rates for the base period can be directly derived from historical data, if such data are available. More often, however, these data are available only for some years of the base period and/or may be of poor quality, especially for small areas. In such circumstances, Booth (2006) says three approaches may be used to derive the requisite fertility, mortality and migration rates for the base period. Some of the approaches presented in this section are similar to the methods presented in section 2.3. The major difference is that in this section the approaches are for estimating the separate components of population change, which are the required inputs into the cohort-component model.

A first approach is to use expectation methods. Under this approach, expectations about future demographic developments can be used to chart the future course of the components of population change. These expectations can be drawn from opinions of experts in the field or from individuals in a survey about their future expected behaviour. Fertility is the component most readily forecast based on expectation methods. For instance in many (fertility) surveys, women are commonly asked about their expected future fertility. Recent innovations of the expectation methods in mortality forecasting are given by Hauser and Willis (2005) who asked questions about the subjective probabilities of future survival chances directly of individuals. Responses to these questions produced remarkable correspondence with the estimated life table survival. The alternative to individual expectation is to use experts' opinion. Booth (2006) reports that the Delphi-based methods forecasts quantitatively each component of population change by bringing together quantitative and qualitative opinions of experts. Booth further notes that the approach is seen as advantageous in developing countries due to its lesser demand of historical data, but the disadvantages of these methods are that they tend to be unsystematic and limited to aggregate forecasting. Age-specific rates are much harder to forecast using the methods.

According to Booth (2006) the second approach that may be used in estimation of demographic rates for the base period is explanation methods. These methods entail estimating future demographic rates by predicting the most likely course of the demographic events in relation to the socio-economic variables prevailing or projected to prevail in future. In theory, therefore, these methods use structural models, whose basic application, strengths and weaknesses was given earlier. The third approach of deriving the components of population change for cohort-component methods is extrapolation methods. These methods estimate the requisite fertility, mortality and migration rates on the basis of the regularity of patterns in the observed data, which are then extended into the future. Use of extrapolation methods is the most common of the approaches for estimation of components of population change for demographic forecasting (Booth, 2006). This is the approach that will be most used in the current study.

Note should, however, be made that the distinction among these three approaches is not at all times as clear-cut as presented here (Booth, 2006). For instance, extrapolation of observed patterns is carried out here after consultation with experts in the field for reasonableness and plausibility. Hence in a way, it combines extrapolation methods with expectation methods.

Once the base period fertility, mortality and migration assumptions are determined, application of cohort-component methods is straightforward. Before the projection process can begin, though, the analyst must determine the starting point (base year) and the starting population (base population) by age and sex. From a methodological point of view the most rewarding way to derive the base population for population projections is to conduct a census of the area (Simpson, Diamond, Tonkin *et* *al.*, 1996). The obvious limitation of this is that, a complete count of an area (irrespective of size) is generally expensive. Simpson, Diamond, Tonkin *et al.* present a less expensive alternative to a complete count of an area. They report that the population of the area can be estimated by updating previous census data for the area by the changes in fertility, mortality and migration that occurred in the area since the census. Thus, this method combines vital registration data and census data to estimate the population for an area at a desired point.

The concept of using vital registration and age distribution data to obtain demographic statistics and population distribution at a point in time is not new. Whilst working on his dissertation in the late 1960s Lee (2004) reports that he was struck by Tony Wrigley's population reconstitution of an English parish, Colyton. He in the process developed, unknowingly at first, the Inverse Projection (IP) technique (Lee, 1974; Lee, 2004). Going by its name, the technique basically inverts the conventional population projection process. That is, in inverse projection annual births, deaths and an age structure counts are used to project sequences of vital rates as well as age distributions. Whereas, in conventional population projection a sequence of vital rates are used to derive counts of births, deaths and age distributions (McCaa and Barbi, 2004). Inverse projection techniques were built on the assumptions that the age structures belong to one parameter families of age specific rates and the population is closed to migration (Lee, 1974; Oeppen, 1993). The method also requires that the registration of vital events be complete. With such stringent assumptions and requirements Lee (2004) was amazed at just how well the method works. He, however, attributed this to the fact that migration and registration effects may be cancelling out (Lee, 1974). Whilst the registration of births and deaths are relatively good in the DSA, there is little information to believe that migration cancels out the effects on population projections of the differential registration of births and deaths that may result from application of inverse projection methods. Migration might systematically distort the results and give implausible estimates (Lee, 1974).

An alternative technique which attempts to address the closed population assumption of inverse projection is the back projection technique (Oeppen, 1993). This technique starts by assuming that the oldest age group at time t-5 is equal to the product of this same age group at time t and the ratio of birth cohorts these two age groups come from. This is contingent, however, on the histories of mortality and migration being the same for these two cohorts. Using the initial estimate of the population in the oldest age group at time t-5, the age distribution at time t and a model life table, the entire age distribution at time t-5 can be estimated using back projection methods (Lee, 1985). Any difference between the estimated 0-4 years age group and the recorded births in the interval (t-5, t), is assigned to net migration and distributed over the cohorts' life, consequently affecting the initially estimated age distribution (Lutz, Goujon, Samir *et al.*, 2007). The estimated age distribution is, therefore, highly dependant on the preliminary oldest age group estimate. This can be a huge problem especially if the source of data for the age distribution at time t is from the census. According to Lee (1985) the population in the oldest age group could be wrong by as much as more than 100 per cent.

In later work Lee (1985) demonstrates that the inverse projection technique still works well even for a population not closed to migration due to the concept of weak ergodicity. He argues that when a population is subjected to a constant regime of vital rates and the number of net migrants was to stay the same for a sufficiently long period prior to the terminal date, the terminal age distribution tends to be independent of the starting age distribution. In his words, the population forgets even its own history. It is also possible to estimate net migration rate if there are true population sizes at two time points and the rate of migration is assumed to be constant in that interval. Oeppen (1993) notes, however, that concept of ergodicity does not apply to the total population size.

As a result Lee (1985) argues that the, simpler to understand and use, inverse projection is to be preferred over back projection, despite it having its own deficiencies. The strength of inverse projection is that it only requires vital registration data to be reliable. It still works well even where census data may be lacking, infrequent or unreliable (McCaa and Barbi, 2004). In this study the technique of inverse projection underlie the methods used to estimate the base population. It also underlies the process of fitting the model for the study area, because the aim of the study is to reconstruct the population distribution of the study area, as robustly as possible. An estimate of the population in base year, 1985, is used to reconstruct the demographic history of the area upto the terminal year of 2005.

2.4 Chapter conclusions

In this chapter a simpler version of the cohort-component method that may be applied in small-area population projections was presented. This simpler version, however, is inadequate for the current study for several reasons. First, it is applicable for projecting the population 5 or 10 years post-census. The target projection horizon in this study is 20 years. Second, the study area experiences a significant level of migration hence use of a method that does not account for this component of population change is likely to produce biased results. Third, the simpler technique can only be used to obtain age-sex projections. In this study, interest is not only demographic projections but epidemiological modelling of the study area as well. Thus, despite the full cohort-component method being complex, data-intensive and time consuming (Smith and Tayman, 2003), it is the chosen method for this study. The underlying methodology in the ASSA model is cohort-component methods.

This study will make use of the data-intensive cohort-component methods since there is sufficient data at several points in time available from a longitudinal data collection system in the study area, which will be used to both parameterise the model as well as test the projection accuracy of the model. The methods are more robust and allow for a more detailed analysis of the study population to be made than is the case for trend extrapolation or ratio methods. Structural models require a range of socio-economic data, whose compilation and analysis was beyond the scope of this study. Hence these models could not as well be applied in this study. Other limitations and problems of trend extrapolation, structural and ratio models which made them inappropriate for this study have been described above.

Despite there being vital events data from a longitudinal surveillance system, national census data for the study area are unreliable. Hence, the common approach of combining census data with vital registration data to estimate the base population could not be applied in this study. Another problem faced in this study was that there was no population for the base year against which the desired base population could be benchmarked. This lack of a benchmark population which could be targeted in deriving the base population offered both opportunities and constraints for this study. The opportunity is that virtually any population size and structure could be used as the base population. The constraint, however, was that the aim of this study was to replicate an already observed population. Clearly not any population size and structure could be able to achieve this. The approach taken here was therefore to benchmark the base population against the 2001 demographic surveillance population. The idea being to estimate the population of base year, 1985, which could have given rise to the 2001 target year population (see chapter 5 for details).

The principles applied here in the estimation of the base population and fitting of the model for the area are those that underlie inverse projection methods developed by Lee (1974) in the 1960s.

3 METHODOLOGY

3.1 Introduction

This study models the demographic and epidemiological dynamics of the area covered by the Africa Centre Demographic Information System (ACDIS) in KwaZulu-Natal province, South Africa. The contextual framework within which this study is undertaken is to investigate the methodological and practical issues of small-area population projections, especially in a population experiencing a major HIV/AIDS epidemic. This chapter reviews the methods and data sources used to model the epidemiological and demographic data of the study area.

3.2 Research design and data sources

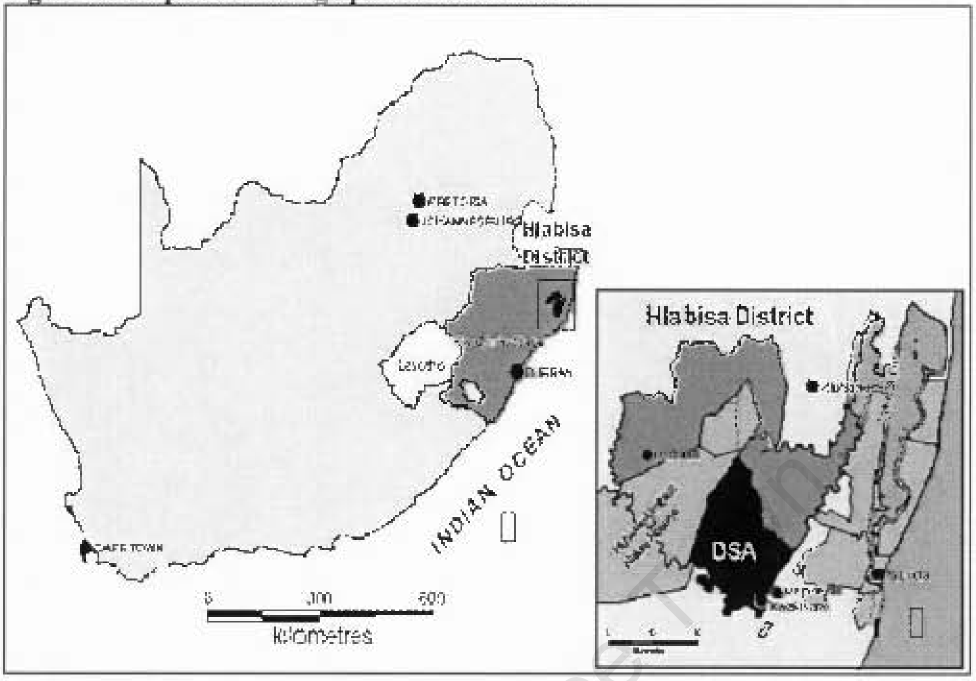
This study uses a non-experimental study design. This is the most appropriate and feasible design given the nature of the study and the wealth of available data. The study is a comparative analysis of the population observed in the ACDIS, the 2001 South African census and the population as modelled using a demographic and epidemiological model. Also presented and discussed in this study, albeit only briefly, are data from the 1996 South African census. Other data required for this study, for example on the HIV/AIDS prevalence in the area, have been derived from various studies and reports.

A description of the study area and the data sources is presented in subsequent sections.

3.3 Profile of the study area

The study area is a demographic surveillance site (DSS) in the Hlabisa health district of KwaZulu-Natal province, South Africa (Figure 3-1). The district is located approximately 250 kilometres north of Durban, the provincial capital (Solarsh, Benzler, Hosegood *et al.*, 2002). Hlabisa health district has four tribal areas. Of interest to this study is the Mpukunyoni tribal area and the urban township of KwaMsane, in which the Demographic Surveillance Site (DSS) is located. The area of coverage of the DSS is predominantly in the rural part of the district and covers only a small portion in the more urbanized township. This demographic surveillance area (DSA) is only 435 square kilometres of the 1, 430 square kilometres of the whole Hlabisa district.

Figure 3-1: Map of the demographic surveillance area



Source: GIS Unit of the African Centre for Health and Population Studies

The population of the study area is predominantly of black Africans. An important characteristic of the study area for projection undertakings is the high level of circular migration in the area (Curtis, Bradshaw and Nojilana, 2002; Ngom, Solarsh, Benzler *et al.*, 2002). Another important characteristic of the area is that it has experienced a dramatic increase in HIV prevalence among women attending antenatal clinics (ANCs) from about 4.2 per cent in 1992 to 36.1 per cent in 2001 (Karim, 2004). The justification for selecting the study area is that, in addition to the availability of comprehensive surveillance data, the features of the area as highlighted above make it ideal to engage with the methodological and practical issues of small area projections. This site is only one of three sites that collect such information in South Africa. Solarsh, Benzler, Hosegood *et al.* (2002) suggest that this area is typical of the rural population in the province. It may also, to some extent, be representative of the rural black population in the rest of South Africa. Hence results of the study to be undertaken here might provide important corroborating data for similar studies in other rural populations in KwaZulu-natal and the rest of South Africa.

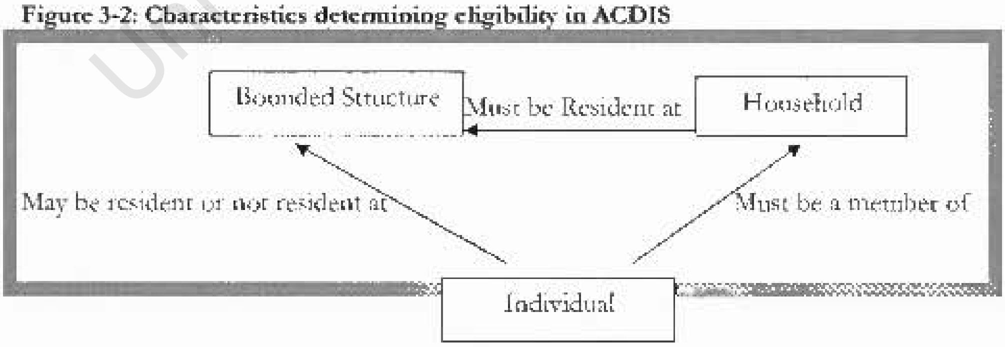
3.4 Design of the Africa Centre Demographic Information System

A brief description of some core concepts and the design of the Hlabisa DSS, more commonly referred to as the Africa Centre Demographic Information System (ACDIS), is imperative here because data from this source are used extensively in this study. Simply put, a Demographic Surveillance System (DSS) is a longitudinal study of a population in a well defined geographical area. The Hlabisa DSA is well delimited by the Umfolozi River to the south, a game reserve to the west, commercial farms on the northern part and the forest reserve to the east. A national road, the N2, separates the DSA from the forest reserve. Such delimitation of the DSA by well-defined physical boundaries makes it easier to ensure that individuals, households and dwelling places that are eligible for inclusion are included and those that are not are excluded.

At the start of a DSS a census is conducted of all people in the defined area. Thereafter, data are collected at regular intervals to capture all persons, households and dwelling places, that joined the DSS after its inception; while key variables and attributes of those registered in the earlier round(s) are updated (Ngom, Solarsh, Benzler *et al.*, 2002). In the case of the ACDIS, information was collected in a 4-monthly cycle from the inception of the ACDIS in January 2000 through to the end of 2001. From 2002, however, the ACDIS has been carried out in 6-monthly cycles. That is, prior to the 2001 census, households in the DSA had experienced between four and six demographic surveillance visits.

There are three important criteria that determine eligibility of individuals for inclusion in the ACDIS. These are best represented diagrammatically, adapted from Hosegood and Timaeus (2006).

Figure 3-2: Characteristics determining eligibility in ACDIS



Taking the thick shaded line to represent the delimited DSA, the central criterion for eligibility in the ACDIS is household membership. For an individual to be included in the surveillance study, he/she must be recognised as a household member of a bounded structure located within the defined DSA. In the ACDIS, a bounded structure is defined as a physical place whose main function is to provide residential accommodation or a selected number of services; while a household is defined as a social group with at least one member residing at a particular physical place (Hosegood and Timæus, 2006). While the ACDIS design allows for an individual to be a member of more than one household within the DSA, they can only be registered to be residing at one bounded structure at a given point in time. The ACDIS further allows for an individual to be registered even if he/she is residing at a bounded structure located outside the DSA provided they are a recognised member of a household residing within the DSA. The term resident population is thus used to refer to household members who reside at a bounded structure located within the DSA, while non-resident population refers to household members residing at a bounded structure outside the DSA.

Another key feature of the ACDIS that is very important to the present study, and to the accurate measurement of exposure to risk more generally, is that of episodes. These are defined as "meaningful and identifiable segments of time started and ended by events" (Ngom, Solarsh, Benzler *et al.*, 2002: 13). The definite start for an individual is the time (date) they were born and the ultimate end is the time they die. In between these two extremes, an episode may end due to several reasons like lost to follow-up, out-migration, or household membership ended. An episode in the ACDIS, therefore, gives the time period that a subject was first under observation until the last time the subject was observed or ceased to be under observation. All episodes therefore, have a start date and an end date, with an identifiable start type and end type; from these demographic events and rates can then be estimated.

3.4.1 Sample size

Demographic and epidemiological data have been collected from all registered households falling within the surveillance area. It is estimated that only about one per cent of all the targeted households in the DSA refuse to participate in the surveillance study (Herbst; Hill; personal communication).

On 1st January 2001 the total registered population in the surveillance study was 89,132 individuals (Hosegood, Vanneste and Timæus, 2004). Of this number about 65,000 were resident within the study area. This study models data that have been collected in the ACDIS from 2000 to 2005. On average each year there are 60,000 to 70,000 persons registered as resident within the DSA and about a further 20,000 registered non-residents. All demographic and epidemiological characteristics occurring in the registered population are stored in a database from which the datasets for this study are derived. This sample size is large enough to compute annual demographic events and rates, the obvious exception being maternal mortality rates. The sample size is also large enough for the ASSA model to run properly. Inputting very small numbers into the model may give nonsensical outputs such as negative numbers of persons.

The demographic surveillance site does not cover the whole of the Hlabisa municipal district, whereas the recorded census population pertains to the entire municipal district. Census data consistent with the DSA, therefore, had to be estimated.

3.5 Mapping of census data onto the DSA

In order to ensure that the enumerated census data is comparable to the DSS data, as exact a proportion as possible of the enumerated census population that falls within the Demographic Surveillance Area (DSA) had to be obtained. This was achieved by making use of Geographical Information System (GIS) maps produced by the Africa Centre for Health and Population Studies. The maps show the exact proportion of each census enumeration area that falls within the DSA (Tanser, personal communication).

Estimation of the population that overlaps with the DSA was straight forward in the case of the 1996 census, but less so for the 2001 census. The enumeration area (EA) codes from Statistics South Africa (StatsSA) changed between the 1996 and the 2001 censuses. Whereas the EA codes from the 1996 census were the same as those listed in the GIS maps for the district, the EAs for the 2001 census were differently coded from those listed on the GIS maps. The StatsSA dataset for the 2001 census provided the data in much bigger areas than in the previous census. On the GIS maps these larger areas were sub-divided into even smaller enumeration areas. Thus, for each StatsSA census enumeration area code there were usually multiple unique codes for each of the subdivisions of the particular census enumeration area given on the GIS maps. For some areas, not the whole of these sub-divisions of the census EAs fell within the DSA. For this study, therefore, to get the proportion of a census EA falling within the DSA, an average proportion of the areas on the GIS maps having the same census EA code was obtained.

The estimated proportion of each EA falling within the DSA was then multiplied by the respective census enumerated population in each age-sex category for that particular EA to get the contribution of that EA to the population distribution for the DSA. The implicit assumption in this estimation is that the population of each EA is uniformly distributed in terms of density. This appears to be a reasonable assumption to make considering that people in the study area, contrary to the norm in most rural areas of Africa, live in homesteads scattered across the surveillance area (Hosegood and Timæus, 2006). Clustered identifiable villages or communities are virtually non-existent in the study area. The assumption can further be justified on account that the percentage of EAs overlapping with areas outside the demographic surveillance area (Table 3-1) were not too many and most had the biggest proportion falling within the DSA. Of the 22 overlapping EAs in the 1996 census only 3 had less than 85 per cent of the area proportion falling within the DSA, whereas the corresponding number in the 2001 census was only 6 EAs. Hence the impact of the bias that may be introduced by this assumption is likely to be minimal.

The census populations mapped onto the DSA is presented in chapter 4.

Table 3-1: Hlabisa District EAs within the DSA and those overlapping

	1996 census		2001 census	
	Number	Per cent	Number	Per cent
EAs Overlapping	22	38.6%	30	39.0%
EAs Completely within	35	61.4%	47	61.0%
Total	57	100.0%	77	100.0%

3.6 Description of ASSA model

The underlying method in the ASSA model is cohort-component methods. The tool utilised in modelling the population dynamics of the study area is the Actuarial Society of South Africa (ASSA2003) model. A detailed description of the history, logic and mechanics of the model is given elsewhere (Dorrington, Johnson and Budlender, 2005; Dorrington, Johnson and Budlender, in preparation). Here only a synopsis of the model design and operation is presented.

Cohort-component models can be grouped into two main categories: deterministic and stochastic models. The conceptual difference between these two models is that deterministic models are based on population averages, whereas stochastic models are based on individual-level simulations. In the former type of model it is assumed that all members of the population have the same characteristics. The model will thus estimate the expected events from the population given the population characteristics. A feature of deterministic models is that they generate the same solutions each time they are run on the basis of the input data assumed. Stochastic models on the other hand treat each variable as a random variable with an associated distribution and mean. As a result stochastic models produce different solutions each time they are run (Johnson, 2004).

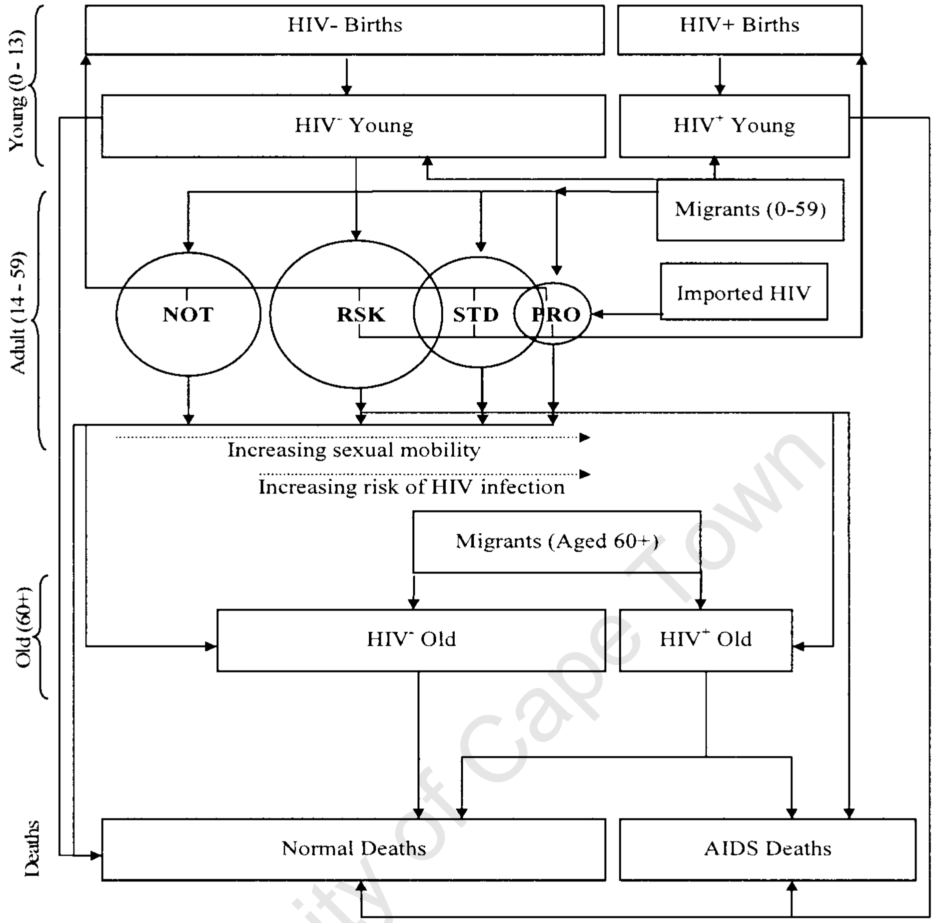
The ASSA demographic and epidemiological model is a deterministic model. The model divides the population into cohorts by age and sex, and identical characteristics are assumed for the cohorts. The actual projections, however, are carried out in single ages. The latest work on the model has involved sensitivity and specificity analysis, to provide a range of possible values for the unique solutions that the model produces given the assumed population characteristics (Johnson, Dorrington and Matthews, 2006).

The first ASSA model was released in 1996. Since then there have been several revisions made to the model necessitated by the need to take into consideration new and/or updated data about the HIV epidemic; need to improve the model fit to data from antenatal HIV surveys; and the need to incorporate new data significant to model calibration (Dorrington, Johnson and Budlender, 2005). These revisions and recalibrations have contributed to making the model more reliable by producing results that more closely match observed reality. The latest release of the model, which is used in this study, is referred to as ASSA2003. The year in the model name reflects the year of the latest ANC HIV data and/or death data used to calibrate the model.

A feature of the ASSA model, which sets it apart from many other models, is that it integrates behavioural, demographic and epidemiological parameters into the projections. Most models typically only incorporate demographic parameters and, to a lesser extent, epidemiological factors. Thus it is more than just a population projection tool. The ability to incorporate HIV prevalence and incidence rates into the projections is important especially for the study area, which is facing a significant burden of HIV/AIDS.

Although projections are carried out in individual ages over a user-defined projection period, when incorporating HIV into the projections the model divides the population into three distinct age groups. The first is the group of the young population (0 to 13 years); the second group is of adults (14 to 59 years); and the third group is the category of the old population (60 years and over). The adult age group is further divided into four risk groups labelled as PRO, STD, RSK, and NOT. The first group is of individuals whose sexual activity matches that of commercial sex workers and their clients. The second risk group (STD) is made up of individuals whose HIV prevalence is identical to persons infected on a regular basis with sexually transmitted infections. This group is also assumed to be similarly infected with sexually transmitted infections as the PRO group. The third risk group, the RSK group, is assumed to have lower sexual activity and about 50 per cent lower condom use than the STD group. The RSK group is, nonetheless, still susceptible to HIV infection given that they are assumed to have at least one new partner per year and also they are assumed to engage in unsafe sex from time to time. The last group, the NOT group, is made up of individuals assumed not to be at risk of being infected with HIV. Even though this group may have lower condom use than the RSK group, they still remain at no risk of HIV infection because they are assumed to have sex only with individuals in the NOT group or not at all. The model assumes that people will remain in the NOT group until they become sexually active in the age range 14 to 24 years. When they become sexually active, they are proportionally moved into the RSK, STD or PRO group. After age 24 years the model assumes no more movement between groups, individuals remain in their allocated risk groups (Dorrington, Johnson and Budlender, 2005). Figure 3-3 illustrates the structure of the model and how it incorporates HIV into the projections.

Figure 3-3: Schematic diagram of the ASSA model



Source: Dorrington, Johnson and Budlender, 2005.

The KwaZulu-Natal ASSA model for black Africans is the model that is being recalibrated to model the population dynamics of the study area. This study aims to contribute to the recalibration of the ASSA model to perform sub-district level projections.

3.7 Ethical considerations

No informed consent from the population of the study area was required given that there was no primary data collected from them. Data for this study comes from the Africa Centre Demographic Information System (ACDIS) database. Since some of the data made available for this study were of a sensitive nature, a data-use agreement has been entered into between the author and the Africa Centre for Health and Population Studies, which stipulated, inter alia, that the analysis and reporting of results will in no way be used to harm or cause prejudice to any person. As such there will be no information presented here that could be used to identify or break the confidentiality the respondents were assured of at the time of data collection. According to the agreement entered into, this report will also be made available to the Africa Centre for health and Population Studies. Every effort has been taken in this study to ensure that all the necessary research ethics are upheld.

3.8 Chapter conclusions

There are some weaknesses with the methods to be applied in this study that may affect the results of the study. Every care was taken, however, to ensure that the impact of these limitations and pitfalls is as minimal as possible. Given the amount of detail desired in the projections and the length of the projection interval (about 20 years) made the other methods inadequate for modelling the population dynamics of the study area. Cohort-component methods and the ASSA model, which combines demographic and epidemiological projections, are indeed the most robust methods to use for this study. It is therefore expected that despite the limitations of this study, the results to be obtained will be reliable and to a large extent a valid representation of the study area.

4 CENSUS AND DSS DATA EVALUATION

4.1 Introduction

Chapter 3 described the methods used to answer the research objectives of this study. The aim of this chapter and the next one is to present and discuss the results of the application of these methods.

This chapter has four main sections. In the next section (4.2) census data relating to the demographic surveillance area (DSA) are presented and discussed. The data are interrogated for possible age and sex errors and other irregularities. Section 4.3 presents a comparison between the unadjusted census population and the adjusted census population for the study area, and discusses the implications of using either of these distributions in population projections. Section 4.4 compares the adjusted census population to the Africa Centre Demographic Information System (ACDIS)[3] data. Finally, section 4.5 compares the unadjusted census population to the ACDIS data. A section on conclusions of the chapter is then presented last.

The 1996 and 2001 census populations mapped onto the DSA (Table 4-1) imply an inter-censal growth rate of 7.14 per cent per annum. This is implausibly high for a population characterised by high out migration and falling fertility. Such a growth rate suggests one or both of the census populations mapped onto the DSA are significantly flawed. A comparison of the census populations and the DSA population points particularly to possible under-enumeration in the 1996 census (Comparison with the 1996 census not shown here, while comparisons with the 2001 census are presented later in this chapter). Results suggest the 1996 census mapped onto the study area is severely flawed. Lack of empirical data from the surveillance area for that particular year made it difficult to interrogate the 1996 census data in greater detail.

The primary objective of this chapter is to investigate whether 2001 census data for the study area can be used in setting the base population in the modelling process of this study. As shown in chapter 2, census data are the basis for base population estimation for many estimation methods.

Advanced data analysis is done using STATA 9 (StataCorp, 2006) computer software. Bivariate and multivariate tables are used to present the results.

[3] The demographic surveillance component of the Hlabisa DSS is formally referred to as the Africa Centre Demographic Information System (ACDIS). In this study whenever possible, the acronym ACDIS will be used when specific reference is being made to this demographic surveillance site. For general discussions the acronym DSS is used.

4.2 Evaluation of census data

The process of estimating the census population that overlaps with the demographic surveillance area was explained in chapter 3. Table 4-1 shows the census population for the area overlapping with the area of coverage of the Africa Centre Demographic Information System, derived from the 1996 and 2001 censuses. In addition to other edits and adjustments, the published census population presents the enumerated population adjusted for the estimated undercount. For purposes of this discussion, the published census population will be referred to as the adjusted census population, while the term unadjusted census population will be used to refer to the census data as enumerated, that is, before adjusting for undercount.

Table 4-1: Census population in five-year age groups and sex, mapped onto DSA

Age group	1996 Census			2001 Census		
	Male	**Female**	**Total**	**Male**	**Female**	**Total**
00-04	3,665	3,725	7,389	4,693	4,694	9,387
05-09	4,143	4,239	8,382	5,404	5,322	10,726
10-14	4,277	4,308	8,585	5,726	5,590	11,316
15-19	3,340	3,626	6,966	5,261	5,537	10,797
20-24	2,185	2,886	5,071	3,037	3,798	6,835
25-29	1,166	2,000	3,167	2,241	3,143	5,384
30-34	980	1,733	2,713	1,690	2,468	4,158
35-39	905	1,579	2,484	1,524	2,212	3,737
40-44	735	1,077	1,812	1,251	2,014	3,265
45-49	616	861	1,477	1,045	1,428	2,473
50-54	380	637	1,016	799	1,134	1,932
55-59	432	734	1,165	611	843	1,455
60-64	281	736	1,017	589	1,090	1,678
65-69	362	645	1,008	383	831	1,214
70-74	222	321	543	371	768	1,139
75-79	151	226	377	183	286	470
80-84	73	133	206	115	235	350
85+	41	132	173	58	158	216
Total	**23,954**	**29,596**	**53,550**	**34,982**	**41,551**	**76,532**

Source: SuperCROSS, 2006, Community profiles tables, Space Time Research Pty Ltd.

Both the 1996 and the 2001 census populations are taken from community profile tables of the SuperCROSS computer database (Space Time Research, 2006). Unlike the 2001 census, the 1996 census population given in the community profiles had a category of age unspecified. The population so classified was redistributed proportionally across all known ages on the assumption that age is equally inaccurately reported at all ages. Even though it is unlikely for non-response to have been the same across all ages, this redistribution is still justified. According to Henry (1976), the error introduced arising from pro rata redistribution is minor if the unspecified category is not significant, as is the case regarding the 1996 census data mapped onto the DSA. The unspecified age category by sex, and for the whole population, was only 0.4 per cent.

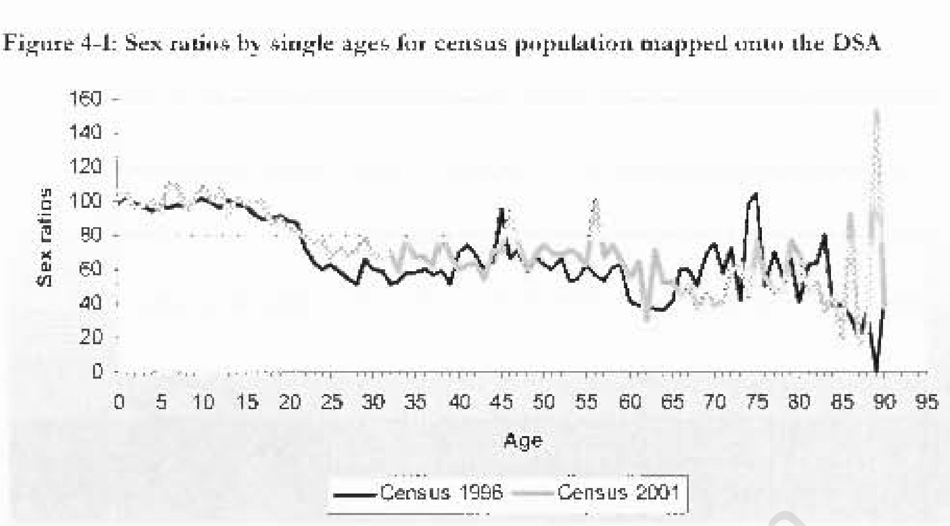
Given that "[N]o census is perfect" (Henry, 1976), the following sections discuss the evaluation exercises performed on the mapped 1996 and 2001 census population distributions for errors that they may contain. Omissions, duplications, misreporting and misclassifications are bound to happen in any census. The effects of these problems are more significant for sub-national areas than the national population. According to Ghosh and Rao (1994), estimates for small areas are likely to display unusually high standard errors owing to their small size.

4.2.1 Interrogation of data for age-sex errors

The data presented in Table 4-1 can be used to investigate the presence of age and sex errors in the data. For instance, the sex ratio distribution may indicate errors in reporting of either sex or both. Sex ratio is defined as the number of males per 100 females (Newell, 1988). The expected trend is for sex ratios to start off at slightly more than 100, as there are usually more male than female births. Evidence from Southern African countries, however suggests that the sex ratio at birth is typically close to one hundred males per one hundred females (Garenne, 2004). Sex ratios typically decline smoothly with increasing age as male mortality tends to be higher than female mortality. A smooth trend in the sex ratios, therefore, is generally indicative of good quality age and sex reporting. An uneven sex ratio distribution in the absence of explanatory factors such as the impact of mortality or migration effects is indicative of age and sex errors.

Figure 4-1 shows that there were significant fluctuations in the sex ratios by age in the two censuses, especially after age 20. The ratios show a sharp drop in the sex ratio from around age 22 in the 1996 census data, indicating that the number of men sharply drops off at this age. The sex ratios stay at very low levels between age 22 and age 40. Migration would account for most of this apparent deficit of men. In addition to being age selective, migration has been shown to be sex selective, especially if the movement is for economic reasons such as to search for employment. Citing Kok et. al. (2003), Wentzel and Tlabela (2006) report that between the periods 1975-1980 and 1992-1996 there were predominantly less female than male internal migrants in South Africa virtually across all ages. These authors also report that migrants were more inclined to be in the age range 15 to 44 years, with the peak between the ages 25 to 29 years.

Figure 4-1: Sex ratios by single ages for census population mapped onto the DSA



An unlikely trend in the sex ratios is the noticeable trough at ages 60-64 in the 1996 census and to a lesser extent about five years later in the 2001 census. Such a trough implies a deficit of men or a surplus of women. It is very unlikely, however, for men to be migrating out in substantial numbers at these ages. What is generally expected instead is for them to be returning at around these ages after retirement (Hill and Queiroz, 2004). It is more likely that women may be exaggerating their ages perhaps so as to qualify for social welfare old age grants.

The sex ratio distributions seen in Figure 4-1 are, therefore, consistent with the expected significant out migration by men in the early twenties to the mid-forties. The trends in the older ages, after age 45 years, on the other hand, fluctuate rather widely and are not consistent with expected trends of a smooth decline. In these ages, however, the effect of having small numbers may be distorting the sex ratio distributions.

There is no strong basis on which to base an argument of significant age-sex errors in the data from the analysis of sex ratios. The expected smooth decline in the sex ratios with age is more applicable in a population closed to migration or a population experiencing negligible migration. For the study area migration is a significant component of the area's population dynamics. Therefore, migration effects are confounding the effect of age-sex errors on the pattern of sex ratios. Attempts were made to interrogate the data for digit preference and age shifting using the standard demographic techniques of the Meyers' and Whipple's indices, but no satisfactory results were obtained. This may be attributable to the fact that the indices are based on assumptions that may not hold for small areas. These indices were designed for analysis of national-level data and their robustness has not yet been established for small areas.

4.3 Comparison of unadjusted to adjusted 2001 census population

In the next two sections the adjusted census population is compared to the unadjusted census population, and each of these distributions to the observed population in the demographic surveillance study to further determine the reliability and accuracy of the census populations mapped onto the DSA.

One error common to all censuses is that of incomplete coverage. While, in some countries and regions the undercount error may be very small, in developing countries it can be a significant source of error in census data. A post-enumeration survey (PES), as the name suggests, is a survey carried out on a sample of the target population in a census immediately following the count. The primary reason that PESs are carried out is the recognition that censuses always suffer from coverage errors. The major objective of a PES is therefore to provide a statistically concrete estimate of the under or over coverage of the census. On the basis of the PES the census count can then be adjusted if the coverage error is significant. To a lesser extent, the PES is also used to evaluate the content of the census, that is, to assess the quality of the information collected in the census count (Statistics South Africa, 2004).

Immediately following the 2001 census, Statistics South Africa carried out a PES, from which it was estimated that the 2001 census undercounted the South African population by about 17.64 per cent. The highest provincial undercount rate (22.51 per cent) was for KwaZulu-Natal province (Statistics South Africa, 2003b). That is, over 1 in 5 people were not counted in the census in the province in which the study area is located.

The estimated undercount rates form the basis for establishing adjustment factors used to adjust the actual census count (Statistics South Africa, 2003b). From the description of the derivation of the adjustment factors by StatsSA, the adjustment factors were not estimated at individual ages and by EA. The weights were instead derived for broad adjustment classes (or stratifying variables) according to province, by EA type, by population group and for broad age groups (0-19, 20-44, and 45 and over). These adjustment classes were then assumed to have a uniform coverage rate, a uniform undercount rate and hence a uniform adjustment factor was applied to each class. Adjustment factors were defined as the reciprocal of one minus the undercount rate. StatsSA reports that the census population was then adjusted by multiplying the actual count by the adjustment factor (Statistics South Africa, 2004).

Since the unweighted populations from the census are not made available by StatsSA due to concerns about confidentiality, the reported census population mapped onto the DSA is adjusted using adjustment factors, presented in Table 4-2 and derived as explained below, to get an estimate of the study area's actual population enumerated in the census. This is, hereafter, referred to as the unadjusted census.

The factors in Table 4-2 were obtained by examining the 10 per cent census sample for the municipalities of Hlabisa and Mtubatuba (municipal codes 535 and 536, respectively). The DSA is located in Hlabisa municipality, but also covers parts of Mtubatuba municipality. The mean weights in the 10 per cent sample and their standard deviations were derived for combinations of the stratifying variables stated above. Attempts were made to get the 100 per cent census data for the study area by EA from StatsSA. The census agency was not helpful in this regard and the issue of confidentiality was given as the main reason for not providing these data.

Table 4-2 shows that the adjustment factors estimated for this study for the study area were highest among the males (1.40) and females (1.32) aged 20-44 years living in urban EAs. The adjustment factors given in Table 4-2 imply undercount rates of about 28.7 per cent and 24.0 per cent respectively for males and females in urban EAs. For the EAs located in the tribal areas of the study area, the enumerated census population is reported to have been undercounted by about 20 per cent for both males and females and across all the age ranges, except for the population in the oldest age groups, which was undercounted by about 18 per cent. These adjustment factors show that the undercount rate is higher in the younger age ranges than in ages 45 years and over among the rural EA population. In contrast, the undercount rate is higher among the adult population 20-44 years old than in the 45 years and over age group, in the urban EAs.

Table 4-2: Adjustment Factors for broad age groups by EA type and sex, census 2001

	Tribal		Urban	
Age group	**Male**	**Female**	**Male**	**Female**
0-19	1.2571	1.2568	1.2342	1.2333
20-44	1.2569	1.2582	1.4021	1.3157
45+	1.2152	1.2132	1.3044	1.1854

With the help of the GIS maps produced by the Africa Centre demographic information system (ACDIS), the reported census population was categorised into rural and urban EA type. The unweighted or unadjusted census population was then reconstructed for the study area from the published census population estimated to be overlapping with the study area by dividing the relevant adjustment factor by the reported population by age and the EA type. The reconstructed unadjusted census population for the study area derived from the above process is presented in Table A- 1.

The composition and structure of the unadjusted census population and the adjusted census population are very similar, resulting from the application of weights that are very similar. There is, however, not enough information to be able to make a judgement on which of the two population distributions would be appropriate for use in population projections. The ACDIS provides a good independent source of data to help in assessing the reliability of the census data. Section 4.4 and section 4.5 demonstrate what effect the application of uniform weights to the broad categories is likely to have on the population distribution of a smaller area. The adjustment factors applied by StatsSA were estimated at a higher geographical unit, the province, but used to adjust the census population down to the smallest unit, an EA.

4.4 Comparison of DSS data to adjusted Census population

In this section of the analysis a comparison is made between the 2001 census population (as adjusted for undercount) that overlaps with the demographic surveillance area and the estimated Africa Centre Demographic Information System (ACDIS) resident population as at 10th October, 2001 (census date).

In terms of population size, analysis of the data shows that the adjusted 2001 census population is greater than the estimated resident population as at 10th October 2001 by about 15.5 per cent (data presented in Table 4-1 and Table 4-3, respectively). Differences are also apparent in terms of structure and distribution.

A comparison of the age distributions (Figure 4-2 and Figure 4-3) shows that there are more people at each age in the 2001 census population mapped onto the DSA than the ACDIS resident population as at 10th October 2001, except for the population below age 10 years and in the old ages (after age 60 years for males and age 50 years for females).

Figure 4-2: Population distribution adjusted Census vs DSS, Males, 2001

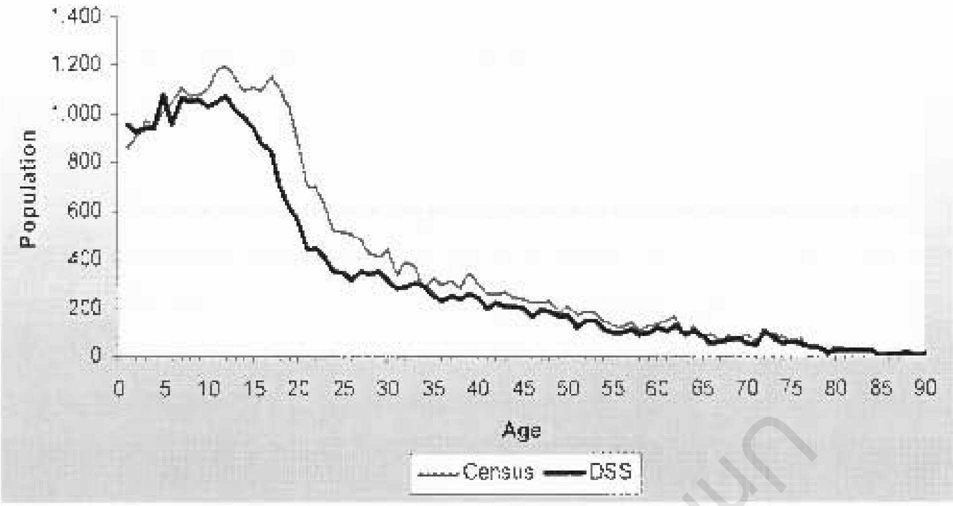
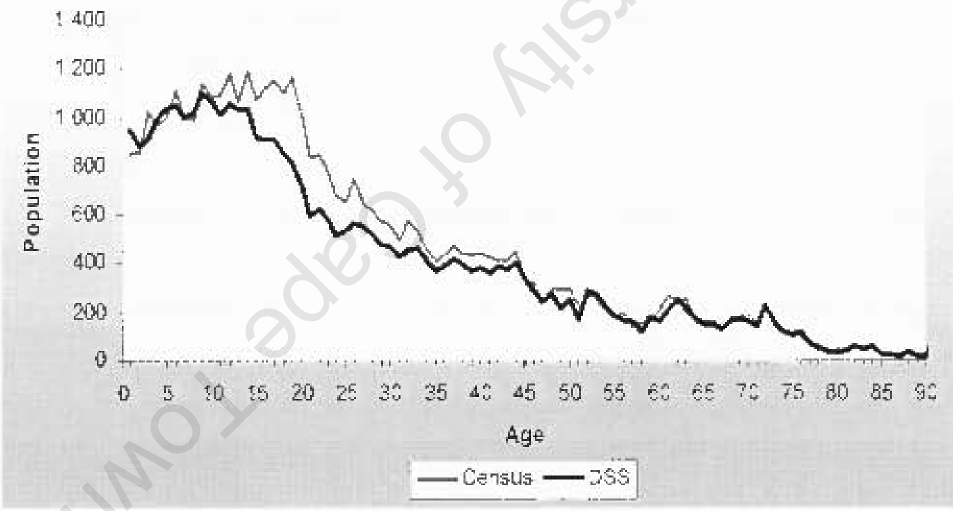


Figure 4-3: Population distribution adjusted Census vs DSS, Females, 2001



An interesting observation is that despite the census having a larger overall population than the resident ACDIS population, the population at the very youngest ages is higher in the ACDIS population than that recorded in the census. This provides evidence of the 2001 census having undercounted the population of children, as is usually the case in developing country censuses. There are significantly more people in the census population than in the ACDIS population just after the teenage ages. The difference is widest in the age range 15 years to about age 30 years. On first impression this may suggest an overcount in the census enumeration. An overcount in the census is, however, highly implausible. It is thus argued that the major reason that the 2001 census population mapped onto the DSA over estimates the population of the study area is because the undercount rate for the study area was most likely significantly lower than the undercount rate (about 25 per cent) used to adjust the census count for the areas in KwaZulu-Natal province.

Many households in the study area by the time of the 2001 census were already under demographic surveillance and therefore, the population of the study area by the time of the census had had extensive experience of being counted. Due to this, the levels of under-enumeration are most likely to be lower than for the rest of KwaZulu Natal. A comparative analysis of the unadjusted census population, the adjusted census population and the resident ACDIS population at census date, in section 4.5 supports this assertion.

Table 4-3: DSS Resident population as at 10th October 2001, Hlabisa

Age	Male	Female	Total
00-04	4,830	4,757	9,587
05-09	5,152	5,228	10,380
10-14	5,056	5,044	10,100
15-19	3,596	4,197	7,793
20-24	1,972	2,851	4,823
25-29	1,659	2,583	4,242
30-34	1,398	2,121	3,519
35-39	1,204	1,950	3,154
40-44	1,022	1,864	2,886
45-49	864	1,261	2,125
50-54	612	1,130	1,742
55-59	506	780	1,286
60-64	495	977	1,472
65-69	300	764	1,064
70-74	329	751	1,080
75-79	159	296	455
80-84	97	211	308
85+	54	186	240
Total	**29,305**	**36,951**	**66,256**

4.5 Comparison of DSS data to unadjusted Census population

One distinctive feature that can be seen from Figure 4-4 and Figure 4-5 is that the unadjusted census population in relation to the ACDIS resident population significantly undercounted the study area's population below age 15 years. At adult ages, however, the unadjusted census population distribution and the resident population are essentially the same; suggesting the unadjusted census population in these ages accurately represents the ACDIS resident population as at census date. The slight differences observed can be attributed to random errors. Similar observations can also be made from Figure 4-6, which shows the ratio of the unadjusted census population to the ACDIS resident population. The figure shows that there was some over-enumeration of the population between age 15 and age 30 years, especially among males. However, an over-enumeration for an African census is not a common occurrence, unlike an under-enumeration. The suggested over-enumeration in the 15-30 years age group was assumed, therefore, to be more an artefact of the data and the reconstruction process than of an actual over-enumeration, as will be discussed later in this paper.

The above analyses provide some evidence to support the evaluation of the 2001 South African census made by the Census Sub Committee (Statistics South Africa, 2006). They reported among other findings, possible underestimation of children under the age of five years; overestimation of the number of people between age 10 and 20; underestimation of men in relation to number of women; and age misstatement among the population in the age range 60 74. The Census Sub-Committee further noted that, the problem of underestimation of children and the differential enumeration of males relative to females is a common feature of developing country censuses.

Figure 4-4: Unadjusted census and resident DSS population distribution, males

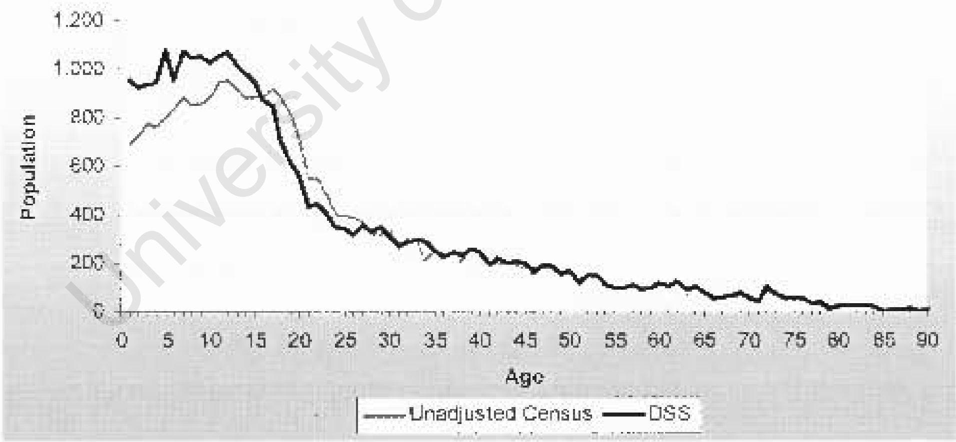


Figure 4-5: Unadjusted census and resident DSS population distribution, females

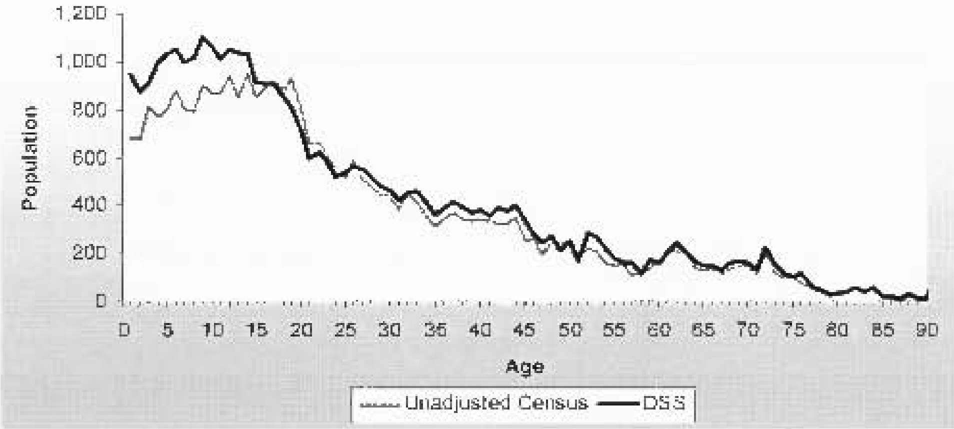


Figure 4-6: Ratio of Unadjusted census to resident DSS population by age and sex

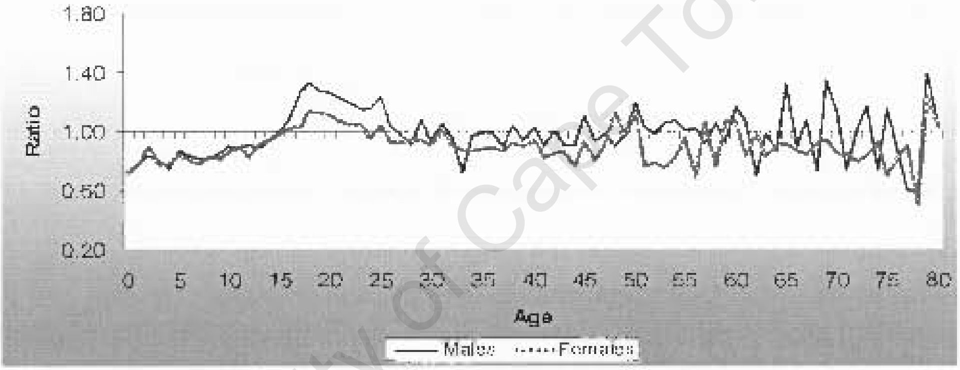
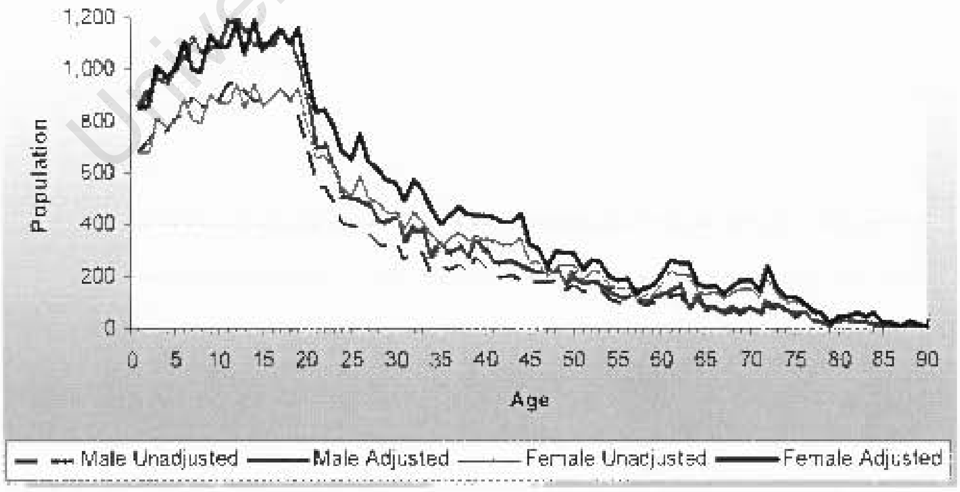


Figure 4-7: Adjusted and Unadjusted population by age and sex, census 2001

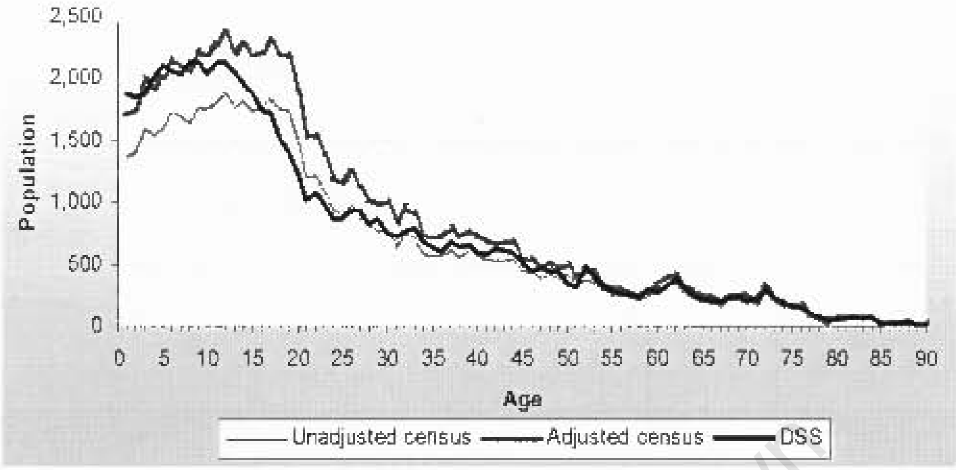


The adjustment factors derived for the study area from the 10 per cent sample suggest that the population of the study area was undercounted on average by about 21 per cent. Relative to the ACDIS resident population, findings here suggest that the adjusted (published) census population mapped onto the DSA may have over-estimated the DSA population. The resident ACDIS population shows apparent under enumeration below age 15 and possible over count between age 15 and 30 in the census, more pronounced among males than females (Figure 4-6). There is also apparent under enumeration of the census male population relative to the female census population as can be seen in Figure 4-7, which shows the number of females among both the adjusted and the unadjusted census population being greater than that of males at most ages.

The estimated unadjusted census population as reconstructed above in relation to the resident ACDIS population at census date suggests that the census under count rate may only have been about 8 per cent compared to about 15.5 estimated from the PES. Most of this under count clearly occurs among children. Given the comprehensive nature in which the ACDIS is conducted, the alternative explanation of the ACDIS significantly underestimating the study area's population is highly unlikely. A comparison of the unadjusted and the adjusted census population to the ACDIS population helps to resolve the problem of which distribution is reliable.

Even though the census population for the DSA was apparently under enumerated, the unadjusted census population offers a more reliable estimate of the study area's population than the adjusted census population. Findings here suggest that the application of uniform weighting factors may have resulted in an inflation of the estimated census population for the DSA. Figure 4-8 shows that, while the adjustment process did provide a good correction for the under estimate of the number of children under age ten, it on the other hand resulted in an over estimation of the population after around age 10 into adult ages. The contention being made here is that applying adjustment factors estimated at a higher geographical unit to a smaller area runs the risk of gravely misrepresenting the population of that area. This situation is further compounded if such a small area has had a good experience of demographic data collection.

Figure 4-8: Distribution of census and resident DSS populations by age, Total



In addition to the effect of the application of the adjustment factors, one other possible explanation for the observed difference between the resident population and the adjusted census population could be attributable to the conception of household membership by the study area's population and its implication on migration estimation. It has been observed that household membership among the population of the study area is not tied to residency in the area (Hosegood and Timæus, 2006). The criteria for household membership are also defined less stringently in the DSS than in the census. In the census the main criteria for membership are to live together, eat together or provide for each other in terms of food and spend most nights in a week at that dwelling unit (Statistics South Africa, 2003a), whereas in the surveillance study membership is largely self defined by the respondent, but it is made clear who is to be considered as a resident or non resident member at the dwelling unit. Due to recall bias of who was a resident and who was a migrant on census night, respondents during the census are more likely to include everyone still considered as a household member. According to Lurie, Harrison, Wilkinson and others:

> "……most censuses are inadequate for estimating the prevalence of migration since they tend not to ask specific questions about migration. Indeed, most temporary migrants in South Africa maintain close links with their rural homesteads. Therefore the census question 'Where do you live?' tends to be interpreted as 'Where is your permanent home?' since most migrants view their migration destination as only a temporary location. Without further probing, important information about migration does not get reported" (Lurie, Harrison, Wilkinson *et al.*, 1997: 20).

This is likely to be a source of problems for the population in the age range 15-30 years, who are more likely to be recent emigrants and therefore, are more likely to still be household members. Unlike in a DSS, the census scope is not able to provide for further probing of people temporarily away and still considered as household members of the area. This may partly explain the inflation of the de facto census population when compared to the recorded resident population as at census night in the DSS.

4.6 Balancing the population

Demographic accounting techniques (or balancing equations) were employed to access the reliability of the ACDIS data. The conventional balancing equation (Hinde, 1998) is given as:

$$P_t = P_0 + B_{0,t} - D_{0,t} + I_{0,t} - E_{0,t}$$

Where, P_t = is the population at time t,

P_0 = the population at time 0;

$B_{0,t}$ = number of births recorded between time 0 and time t;

$D_{0,t}$ = the number of deaths recorded between time 0 and time t,

$I_{0,t}$ = number of in migrants between time 0 and time t

$E_{0,t}$ = number of out-migrants between time 0 and time t

For the balancing equation fitted in this study an additional component had to be added to this conventional equation to take into account the unique features of the ACDIS surveillance study. In the surveillance study, decrements to the recorded resident population will not only occur from deaths or out migration. Some of the registered population simply end their household membership and hence are not included in the surveillance study. Other residency end types that have to be accounted for in balancing the equation includes, household dissolution, household destruction and individuals assigned the not applicable category. Thus a final term to the balancing equation that accounts for all these other event types was added (referred to as 'other losses').

The equation fitted was:

$$P_t = P_0 + B_0 - D_0 + I_0 - E_0 - O_0$$

Where, P_t = is the population at time t,

P_0 = the population at the start of time 0;

B_0 = number of births recorded at time 0;

D_0 = the number of deaths recorded at time 0;

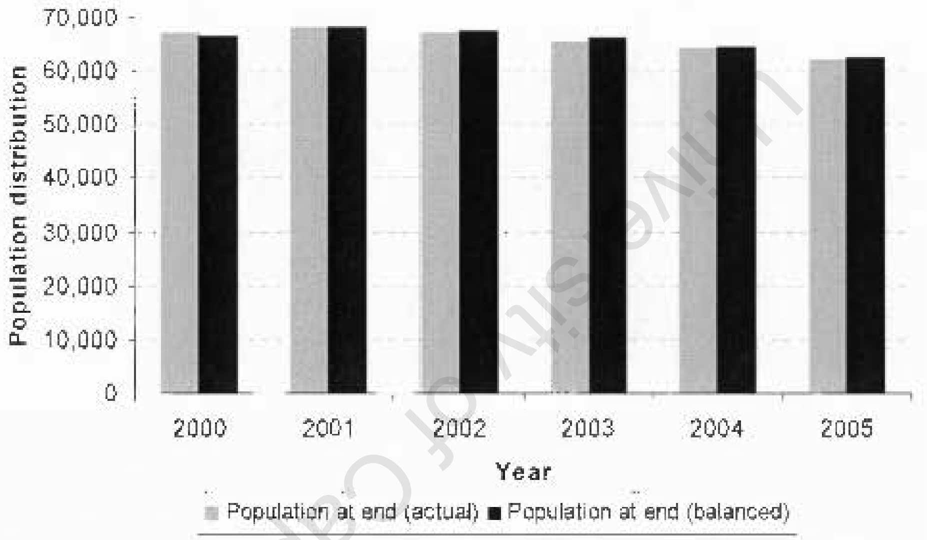
I_0 = number of in-migrants at time 0

E_0 = number of out-migrants at time 0

O_0 = the population ending their residency and membership at time 0 for reasons other than death or out-migration.

Figure 4-9 compares the recorded total ACDIS resident population and the population derived through demographic accounting. The input data for the figure appear in the appendix (Table A- 8 and Table A- 9).

Figure 4-9: Comparison of the recorded and the balanced population for the DSA



An inspection of the percentage differences between the balanced and the observed population indicates that the difference is less than one per cent in all cases except for the year 2003 for the total population. Among males the percentage differences are less than one per cent in all years, while among females there are greater wide fluctuations in the differences, with the percentage differences greater than one per cent in the year 2000 and 2003 (Table A- 8 and Table A- 9).

Where the population distribution and vital events have been accurately recorded there should be zero percentage differences between the balanced and the actual population for each year. There are a number of explanations that could account for the observed discrepancies in the balancing process here. The results suggest that there could be possibly under reporting of births and deaths. Another possible source of problems in the balancing process is the recorded migrants. There is a tendency among respondents in the surveillance study to back-date the start of residency in the household. More recent migrants are also more likely to be reported as still resident in the household. There is also the problem of violating the principle of correspondence in that resident mothers who give birth outside the surveillance area are captured in the study but their babies may not. The converse is also possible, non-resident women are captured in the surveillance, but the babies they bear whilst still non-resident may not be captured. Another potential source of in-balance could be age mis-reporting, as suggested by the high percentage differences in the old age groups. These age groups are usually associated with age misreporting in African populations. Incomplete reporting, as well as mis-reporting, of vital events is therefore likely to be the major source of errors observed in the findings here. This is so because the surveillance study depends largely on self reporting by the respondents.

Detailed investigation of the ACDIS data for these potential sources of error and bias was beyond the scope of this study. It is therefore difficult to categorically state the sources of problems in the balancing process. However, the differences between the recorded and the balanced population are very small, further augmenting the argument that the ACDIS data are highly robust and reliable, and hence the justification to use these in the population modelling process.

4.7 Chapter conclusions

This chapter has compared three different estimates of the population for the study area as at 10th October 2001. The resident ACDIS population estimate as at the census date is assumed to be highly reliable given the comprehensive methodology and scope of this undertaking. It has also been demonstrated in this chapter that the census data, as is usually the case, contains some errors.

A comparison of the unadjusted and the adjusted census population to the DSS population revealed that typical errors such as the under enumeration of children and the under enumeration of males relative to females are apparent in the census population mapped onto the DSA.

As demonstrated above, the adjustment process more accurately corrected for the typical under count of children, but was inappropriate for the adults. There is, however, need for careful investigation of why children are usually undercounted in African censuses. This is certainly an area for further research because even for the study area, whose population is particularly compliant vis-à-vis being counted, there is still evidence of undercounting children.

There is need for careful application of adjustment factors estimated at a higher geographical unit when attempting to correct for coverage errors for a small area, as the coverage rate in the latter may be significantly different from the larger unit at which these factors are estimated. This is likely to result in an over estimation of the small area population as was apparently the case for the study area of this study.

These findings motivated the use of ACDIS data to calibrate and fit the model, as it is demonstrated to be more accurate and reliable than either the adjusted or the unadjusted 2001 census population. The modelling process discussed in the next chapter will attempt to replicate, as closely as possible, the recorded resident ACDIS population for the years 2000 to 2005.

What lends credibility to the findings here is the fact that, as would be expected, the estimated unadjusted census population is less than the ACDIS population at most ages. The only exception is among the ages between 15 and 25 years, more pronounced among males than females, where the unadjusted census population is slightly more than the ACDIS population. The patterns observed in this age range may, however, be attributed more to estimation errors than actual patterns in the data. It is only when the ACDIS population is compared to the adjusted census population that the later suggest a huge overcount in the middle ages. Confidence can be place in the findings here because the unadjusted census provides strong evidence of the expected undercount among children in the census. The implication of overcount in the adjusted census is indicative of definite errors in these data. Possible reasons and/or sources for these errors have been discussed in this chapter.

5 MODEL FITTING AND CALIBRATION

5.1 Introduction

This chapter presents results and a discussion of the calibration and fitting of the ASSA model to the study area. In this chapter, the projected or modelled population is discussed in relation to the Africa Centre Demographic Information System (ACDIS) population.

The version of the ASSA model chosen for use in this study is ASSA2003 'lite' (Dorrington, Johnson and Budlender, 2005). The 'lite' version, unlike the ASSA full model, is calibrated to perform demographic and epidemiological projections on a homogenous population with respect to population group (or race). The 'lite' version models the population as a whole, while the full model computes projections separately for each of the four classifications of the population, namely black African, Indian/Asian, Coloured and White. These separate projections are then aggregated to produce projections for the whole population (Dorrington, Johnson and Budlender, 2005). Use of the 'lite' version here is justified because the study area is predominantly of Black/African race group. Over 95 per cent of the population of the area is of this group (Curtis, Bradshaw and Nojilana, 2002), hence it suffices to treat the population as homogenous vis-à-vis race.

The model fitting process involved initially extracting the parameters for Africans from the full model for KwaZulu-Natal into the 'lite' model. All the demographic and epidemiological parameters and assumptions in the full model for Africans were copied into the 'lite' model, such that the latter model reproduced results from the former model.

This study can be categorised into two main stages. The first stage was to establish a benchmark population distribution for the study area. This was the subject matter of chapter 4, in which the reliability and accuracy of census and ACDIS data was discussed. It was established in that chapter that the enumerated 2001 census data mapped onto the study area are not reliable, hence could not be used as a benchmark. Even the estimated unadjusted 2001 census population was seen not to quite reliably represent the study area. ACDIS data were, therefore, taken as the benchmark population. The second stage of this study, which is the focus of this chapter, is to calibrate the model to reproduce, in the first instance, the 2001 ACDIS population as closely as possible; since this was the population distribution investigated in detail in chapter 4 for its reliability as a benchmark population. The model will then be extended to the other years for which ACDIS data are available for further investigation of its robustness.

The projected population to be derived from the model is the equivalent of the resident mid-year population in the DSA. It projects the population distribution that would be in the DSA given a set of fertility, mortality, migration and epidemiological assumptions over time. Section 5.2 discusses how these assumptions were estimated and/or set in the model. Section 5.3 will then present and discuss the reasonableness of key findings from the model. Finally, section 5.4 will present some conclusions from this chapter.

5.2 Model assumptions

5.2.1 Fertility assumptions

5.2.1.1 Fertility estimation for 2000 and beyond

The fertility rates used to calibrate the model for the period 2000 to 2004 are taken from Moultrie, Hill, Hosegood and Herbst (in preparation). These fertility rates were derived on the basis of reported births only, leaving out data from maternity histories. While maternity histories will capture a birth and thus allow it to be included in the numerator, the mother may not be living in the study area at the time of the birth and hence, the principle of correspondence may be violated. This is likely to result in the over estimation of fertility rates.

A second consideration taken into account in the estimation of these fertility rates was the exclusion of births to mothers who were not resident in the DSA at the time the birth occurred. The likely effect of these considerations is to reduce estimates of fertility in earlier periods, and thus make fertility decline appear not so rapid.

5.2.1.2 Fertility estimation before 2000

The fertility rates needed as inputs into the model for years prior to 2000 are based on work by Moultrie (in preparation), where he looks at the maternity histories of women aged 15-49 at the start of the DSS in 2000. This allowed for the derivation of fertility rates for the period, 1990-1999, preceding DSS data collection.

Retrospective data on fertility were only collected for women aged 15-49 years at the inception of the study. This means that estimates of fertility are increasingly censored with each year preceding the baseline investigation. As such in order to estimate the fertility of women aged 45-49 years five years before the baseline (that is, in

1995) retrospective reporting of fertility data for women aged 50-54 years at the baseline would have been required. Evidently it is the estimates of older-age fertility that are censored first. It is a fortunate coincidence that fertility in the older age groups is both much less significant than fertility in younger age groups, and hence also that absolute changes in fertility in these age groups are less dramatic than those in younger age groups. Hence, in order to derive a longer series of fertility estimates from the baseline study, the fertility rate in the 45-49 years age group for years before 1996 was constrained to be the average fertility rate in that age group in the years 1997-1999 (10 children per 1000 women); while fertility estimates in the 40-44 years age group before 1993 were similarly constrained to be the average fertility rate of women aged 40-44 years in the years 1994-1996 (54 children per 1000 women). If fertility has been falling in this population, as a result of both adjustments the estimated total fertility rates certainly will be slightly under- rather than over-estimated.

This method allows the ready determination of fertility for a decade before the baseline (i.e from 1990 onwards), without having to concern oneself with the estimation of fertility among women aged 35-39 years from truncated data. Accordingly, a different approach has to be adopted to derive estimates of fertility for the five years from 1985 to 1990. An exponential curve was fitted to the TFRs for the period 1990 to 1999, and then extrapolated backwards to 1985.

In order to derive age-specific fertility rates for the projection horizon, 1985 to 2004, first standard fertility distributions for the projection horizon were obtained by interpolating between the standard schedules for Africans in KwaZulu-Natal in 1996 and for the DSA in 2004. The derived fertility schedule for each year was then multiplied by the respective TFR to get fertility rates in five-year age groups for the reproductive age range, 15 to 49 years. After obtaining this distribution of fertility rates the second step was to estimate fertility rates in single ages. This was achieved by applying the Beers Modified interpolation formula (Shryock and Siegel, 1976). The Beers modified formula was preferred over the Beers ordinary formula because the former not only interpolates between the given values but also does some smoothing or graduation of the values, (Shryock and Siegel, 1976). The obtained distribution of age-specific fertility rates in single ages for the period 1985 to 2004 was then imported into the ASSA model for projection of the DSA population.

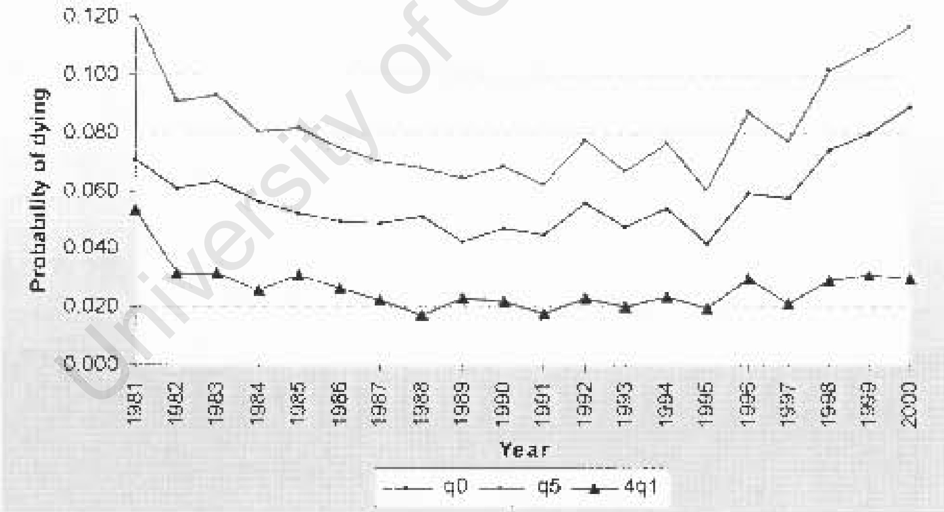
The age distributions for both the post-2000 and the before-2000 fertility rates are almost invariant. This is expected as fertility distributions change very slowly over time. The reliability and accuracy of these fertility estimates will however, be assessed by how well the model, using these rates in combination with the other input parameters, reproduces the study area's population age-sex structure. The total fertility rates derived here are illustrated in Figure 5-10, which compares the projected to the observed fertility rates.

5.2.2 Mortality assumptions

5.2.2.1 Child mortality estimation

The initial child mortality rates for the period from 1981 to 2001 for the study area were obtained from Nannan, Bradshaw, Laubscher *et al.* (2005). These estimates are all cause mortality rates. The model, however, requires as input non-AIDS mortality rates. To get the non-AIDS mortality rates for the projection horizon, the all-cause child mortality rates were analysed for trends and patterns. Figure 5-1 shows a generally consistent pattern of declining all-cause mortality up to about the mid-1980s to the late 1980s, when mortality rates generally show an upward trend. Such a change in mortality rates has been attributed to the impact of HIV (Karim and Karim, 2002).

Figure 5-1: All-cause combined child mortality rates by year



Source: Nannan, Bradshaw, Laubscher *et al* (2005).

HIV in South Africa, as in most Southern African countries, is believed to have started to spread around the early to mid 1980s. The first reported case of HIV in the country was in 1982, but up until about 1990 there was relatively low HIV infection among the general population (Karim and Karim, 2002; Karim, 2004). However, the evolution of the South African epidemic after this initial slow start has been explosive (Karim and Karim, 1999).

It has further been reported by Hosegood, Vanneste and Timaeus (2004) citing Swanevelder, Kustner and Middelkoop (1998) that the epidemic in KwaZulu-Natal initially had a lead over other South African provinces of about 2 to 3 years. In this study, therefore, it is assumed taking into account the observed trends in child mortality rates that AIDS mortality prior to about 1985 was insignificant. Analyses generally showed declining trends in child mortality up to around 1985 and it is further assumed that in the absence of HIV/AIDS the observed trends of declining child mortality rates would have continued into the future. The observed declining trends in child mortality rates are therefore then extrapolated forward from 1985 to 2005 by fitting the following logistic curve to the child mortality rates for both sexes combined (q0, 1q4 and q5).

$$f(t) = \left(a \cdot e^{-b \cdot t} + c\right)^{-1}$$

Where, a, b and c are constants, and f(t) is a function of child mortality rate at time t. This equation was solved using maximum likelihood regression equations. The constants b and c were derived by fitting the points:

$(f(t))^{\wedge-1}$ and $(f(t+2))^{\wedge-1}$ to the child mortality rates for which the declining trend was being assumed will continue. Once b and c were estimated, the constant a was simply obtained by a transformation of the above logistic curve to the form:

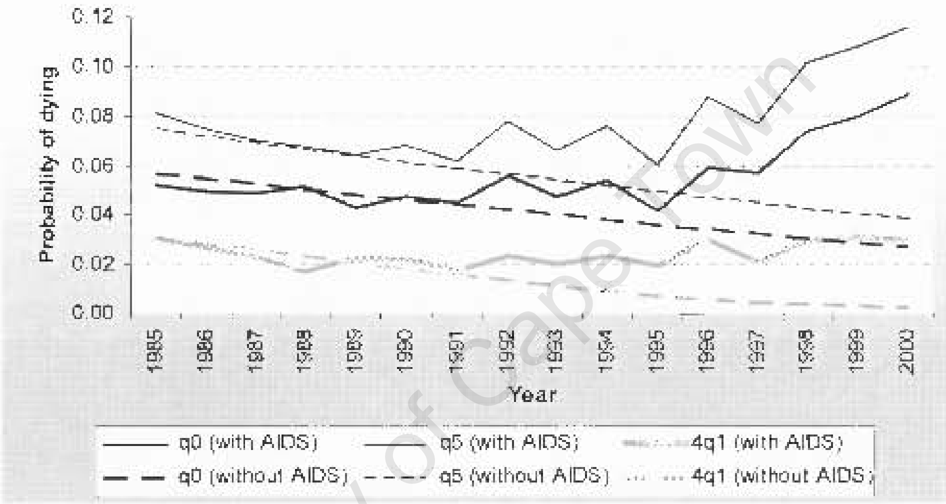
$a=[(f(t))^{\wedge-1}-c]/e^{\wedge-bt}$

The logistic curve was preferred over other extrapolation models like the exponential model because though it is assumed that the observed child mortality would have continued to decline, realistically mortality could not fall to zero and cannot go beyond a certain maximum. It is more likely to follow an s-shape implied in a logistic curve (Hinde, 1998).

The logistic curve was fitted to q0, 4q1 and q5 for the combined male and female mortality. Using the ratios implied by the child mortality rates derived by Garrib, Jaffar, Knight *et al.* (2006), for the period 2000 – 2002 for the DSA, the fitted mortality rates were then separated into male and female by assuming that the ratio of male and female mortality is respectively, 1.03 and 0.95, to the total estimated child mortality.

Figure 5-2 combines the all-cause mortality taken from Nannan, Bradshaw, Laubscher *et al* (2005) illustrated above and the estimated mortality trends that may have prevailed had there been no emergence of the HIV pandemic and assuming the observed mortality trends prior to 1985 were to continue. Figure 5-2 shows how child mortality would have been significantly lower had there been no AIDS impact. The figure also highlights the importance of having a model like the ASSA model which starts from a scenario of no-AIDS, but then incorporates AIDS mortality into the projections according to the epidemic's development. The reliability of projections that do not explicitly take into account the impact of AIDS for a population that has been severely affected is seriously called into question.

Figure 5-2: Estimated child mortality rates with- and without-AIDS



Karup-King graduation techniques (Siegel and Swanson, 2004) were then applied to the under-five mortality rates to get mortality rates in single ages between age 1 and 4 years. Results of this process together with the earlier derived mortality rates at age zero are presented in Table A- 3. These are the rates which were then imported into the model as the non-AIDS child mortality rates for the projection horizon, 1985 to 2001.

5.2.2.2 Adult mortality estimation

Due to time constraints the adult non-AIDS mortality rates required as inputs into the model could not be independently derived for the study area. The model is thus calibrated with the non-AIDS child mortality rates estimated above and the adult mortality assumptions of the provincial model for Africans. It was held that adult mortality in the study area would not be substantially different from that in the province to seriously affect the model results. This seemed to be highly plausible as research findings show HIV prevalence in the district to be very comparable to provincial prevalence (Karim and Karim, 2002, Bradshaw, Nannan, Laubscher *et al.*, 2004; Department of Health, 2006), and among adults AIDS is a major cause of death (Hosegood, Vanneste and Timaeus, 2004). The rationale for using provincial adult mortality rates was, therefore, that adult mortality profiles of the study area are not very different from those of the province. However, as Bradshaw, Nannan, Laubscher *et al.* (2004) argue, this does not excuse the need for a careful comparison of mortality rates by age in the study area and the province to understand the distinctions. This could not be done for this particular exercise for reasons already alluded to earlier, but also that this decision to use provincial rates in fitting the model is unlikely to significantly affect the outcome from the exercise.

5.2.3 Migration assumptions

In the DSS databases each individual has an observation episode with the date the observation started (startdate) and the date last observed (enddate). Each individual is further classified into a specific category at the start (starttype) and their status at the date last observed (endtype). In terms of migration status, all individual residency statuses are left-censored by the startdate. That is, nothing is known before the start of the DSS or before an individual was first observed. It is for this reason that migration events can only be calculated from the DSS data for the period 2000 (when the DSS started) to 2005 (the latest year for which DSS data were available for this study).

Migration rates are estimated as at mid-year from the residency endtype and residency starttype by the date these events occurred. All individuals moving into the DSA from outside the area after the mid-year of a previous year to the mid-year of the year of interest are added to the in-migration events for that year. For instance, all individuals who in-migrate into the DSA from outside the area after mid-year 1999 to mid-year 2000 are added to the in-migration events for the year 2000. Similarly, all individuals moving out of the DSA after the mid-year of a previous year to the mid-year of the year of interest make up the out-migration events for that year.

Given the phenomenon of circular migration that characterises the study area and the need to estimate migration rates on an annual basis, multiple movers are likely to inflate the migration rates. Migration events are summed irrespective of whether or not it is the same individual making the move. The potential impact of this problem is, however, mitigated by the use of net migration rates. That is, circular migration between the DSA and an external area is likely to result in the cancelling out of migration events. A multiple mover will add to both in migration events and out-migration events.

Figure 5-3: Smoothed migration rates per year by age and sex, Males

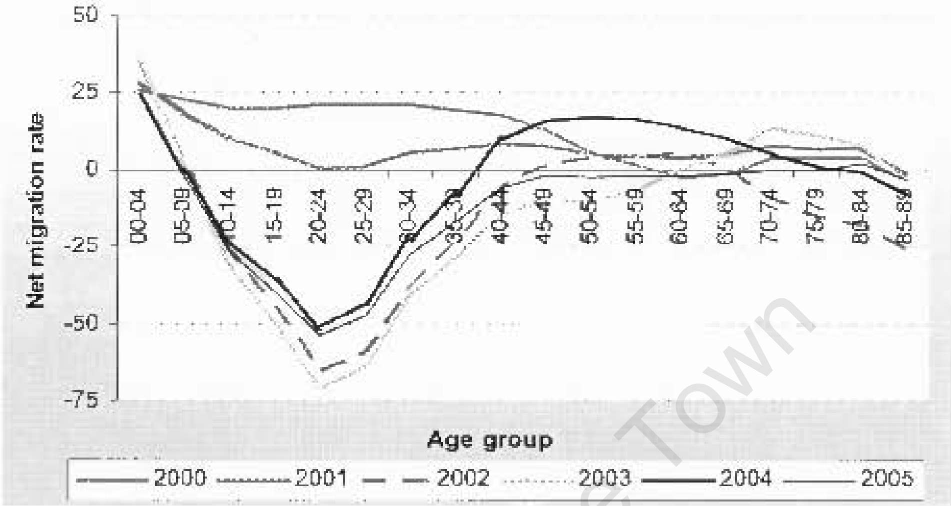
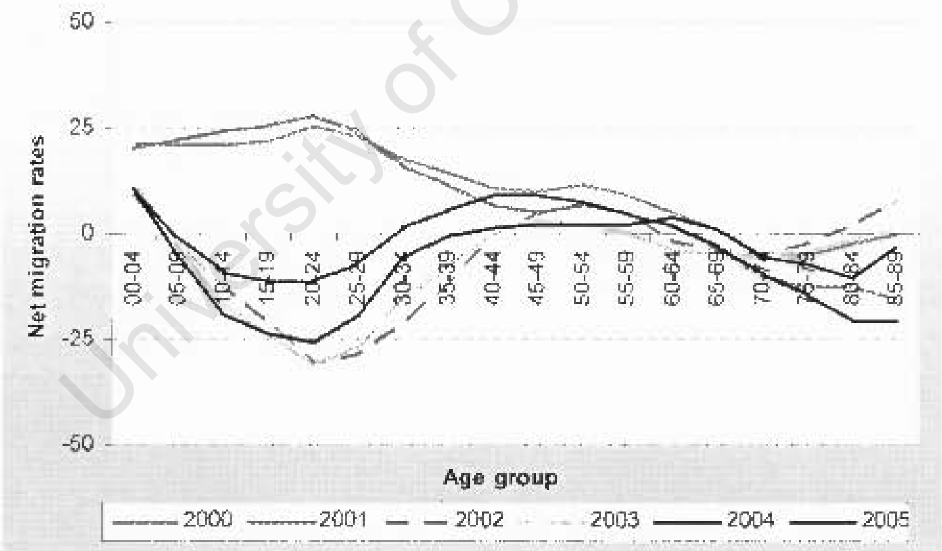


Figure 5-4: Smoothed migration rates per year by age and sex, Females

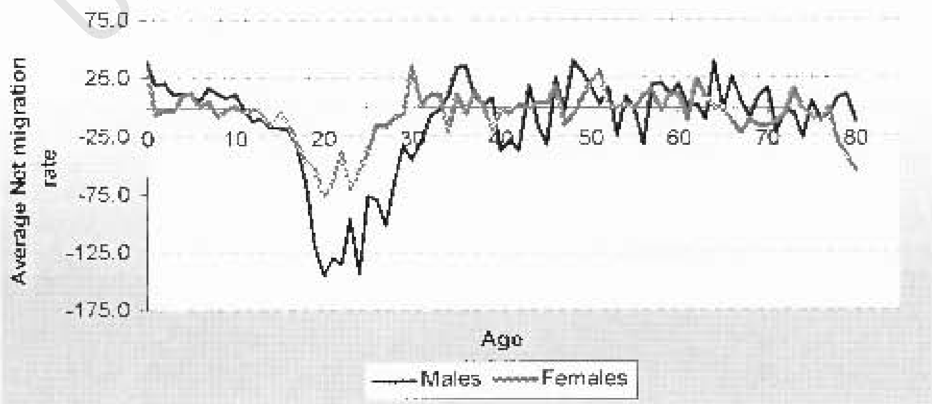


Migration events were estimated initially for the period 2000 to 2005, from the available DSS data. As can be seen in Figure 5-3 (males) and Figure 5-4 (females), other than for the years 2000 and 2001, there is a consistent pattern of net out migration from the study area. The observed pattern for the year 2000 is attributable to insufficient data to estimate migration rates at mid-year since data collection in the DSS started only on

1st January 2000 and thus all residencies are left-censored by this date. Migration rates had to be estimated at mid-year so as to be consistent with the design of the model in that flow items have to refer to a period from 30th June the previous year to 1st July in the current year.

With the exception of the 2000 and the 2001 distributions, it can also be seen that for both males and females the migration events peak around the age range 20-24 years. These patterns are typical of out-migration motivated by economic reasons. Despite such similar trends in migration patterns between men and women in the study area the rates of migration are different in the peak age ranges. This is repeated in Figure 5-5, which shows that on average for the period 2002 to 2005 the rate of male migration per 1000 resident mid-year population is about twice the corresponding rate for females in the peak migration age group of 20-24. As would be expected, migration rates among young children do not vary significantly between the sexes. The expectation is for these to tend to migrate together with their parents. Seeing that the migration rates peak in the mid to late 20s and is for out-migration, may contradict with this expectation given that children are observed to have net in-migration rates. However, this may be explained by the observation that many of the people leave the study area to seek for employment, and so tend to be young, single and in most probabilities without children. The trough of net in migration around 50-55 years may be associated with people coming back to the study area after retirement, losing their job or sick (and wanting to die in their ancestral home). This is the group that is most likely accounting for the net in-migration among children being observed. That is, these are the initial childless out-migrants who are now parents and migrating into the area with their children, as per expectation.

Figure 5-5: Average migration rates per 1000 population by age and sex, 2002-2005



The model was thus fitted with the derived migration rates for the period 2002 to 2005. Migration events for the other years in the projection horizon that is, from 1985 to 2001 were estimated by an iterative process using the ASSA model. An initial set of number of migrants was estimated using extrapolation techniques. These were used to run the model forward from 1985 to 2005. Migration rates were then estimated for each year and compared to the rates derived from the DSS data. If the rates were not comparable the migration assumptions in the model was reset and the model run again. New rates were in turn derived and compared to the rates from the empirical data. This iterative process was continued until reasonable consistency in the migration rates between those from the empirical data and those derived from the model was achieved. The ultimate migration rates settled for from the iterative process are presented in Figure 5-6 and Figure 5-7 for males and females respectively. The model is then calibrated and run using these rates.

Figure 5-6: Migration rates used in model calibration, Males

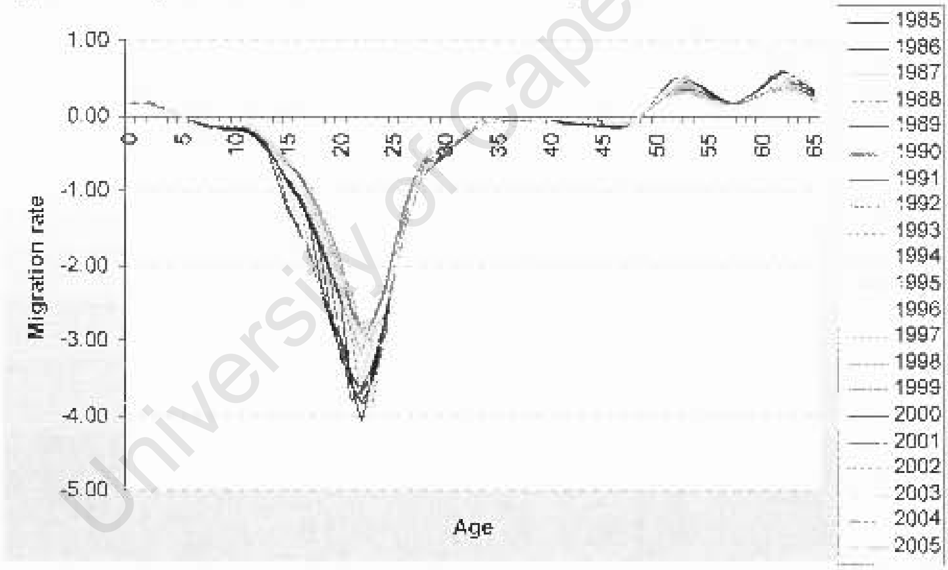
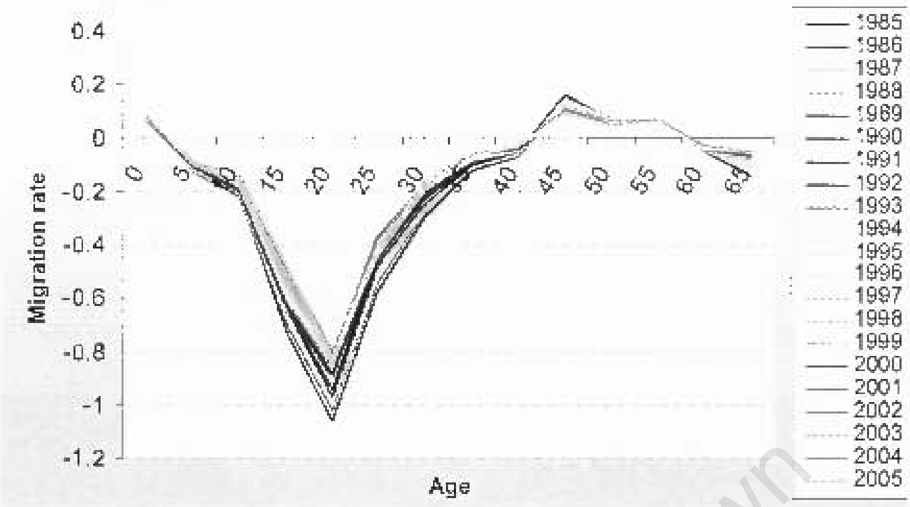


Figure 5-7: Migration rates used in model calibration, Females



5.2.4 Epidemiological assumptions

5.2.4.1 Initial Condom Usage

Condom usage is estimated for the year 2005 for males from the Men's General Health dataset of the DSS. To get the 'average initial condom usage' in the ASSA model, that is condom usage for the base year, a process of backward and forward projections was followed. The model is run forward from 1985 to 2005. Then the projected condom usage in 2005 for the RSK group is compared to the estimated condom usage for 2005 from the DSS data. If these two estimates are not identical the initial condom usage estimate in the model is adjusted appropriately and the model run forward again. Again the obtained condom usage for the RSK group is compared to the estimated condom usage for 2005. This backward and forward projection is continued until the projected condom usage for 2005 closely matches the estimated condom usage from the DSS data. This procedure was followed successively at each age in the reproductive period 15 to 49 years. In this way the required initial condom usage in the model that pertains to the study area by individual ages, essential for HIV incidence and prevalence projection, is estimated (Table A- 12). A more robust way of estimating the condom usage would have been to iteratively adjust the base period and current period condom usages from the model and the DSS, respectively. There was, however, no data on whose basis the current period condom usage could be adjusted. The position taken therefore was to privilege the DSS data on the basis of its robustness. Condom usage at baseline was as a result obtained, as explained above, by trial and error until the usage in 2005 passes through the recorded DSS condom usage.

5.2.4.2 HIV prevalence

The antenatal HIV prevalence data used to calibrate the model pertain to prevalence estimates for the Hlabisa district for the period 1992 to 1999 for women aged 15-49 years attending antenatal clinics (Coleman and Wilkinson, 1997; Wilkinson, Karim, Williams *et al.*, 2000; Karim and Karim, 2002), while for the period between 2000 and 2004 they relate to KwaZulu Natal provincial prevalence rates for women attending antenatal clinics (ANC) in the province (Department of Health, 2004). Given the inadequacy of HIV data, the HIV prevalence curve is fitted to a combination of the DSA specific data and to the provincial estimates. The levels of HIV prevalence rates among antenatal attendees in the sub-district and the province, as argued earlier, are very similar and hence it comes as no surprise that the levels and trends in the area-specific prevalence rates and the provincial rates are very consistent (Figure 5-11).

Given that this is a rural area, it is assumed that use of private health clinics for antenatal purposes is negligible. When using ANC data from public clinics to calibrate the model, the ASSA model adjusts these data by a factor referred to as 'ANC adjustment factors'. Given the assumption above, these factors were set equal to one at each of the five-year age groups in the child bearing years. That is, the public-private bias was removed. The use of private clinics for antenatal care was assumed not to be significantly biasing the model estimations which are based on ANC data from public clinics. Through a process of trial and error the proportions in each risk group that were able to reproduce the observed antenatal HIV prevalence for the study area were: 1.2 per cent (PRO), 40 per cent (STD), and 42 per cent (RSK) (Table A- 10). The model was therefore run using these parameters. For a description of the risk groups refer to section 3.6.

All the other assumptions for projecting the HIV prevalence are set to the KwaZulu Natal ASSA provincial model default assumptions (Table A- 11).

5.2.5 Base population estimation

The base population was obtained in a two step procedure. The ASSA model was calibrated as explained above. To get the initial base population, the model was run from base-year 1985 to target-year 2001 on the assumption that there is zero migration and using an arbitrarily set base population. Survival factors from base-year 1985 to target-year 2001 were then estimated by dividing the projected population aged x+16 in 2001 by the population aged x in 1985. The age group 90 years and over in 2001 was divided by the population aged 73 years and over in 1985. Survival factors could only be estimated in this manner for the population aged 16-years and over in 2001. Those below 16 years were not yet born in 1985. The initial base population was then obtained by dividing the recorded resident DSS population aged 16 and over by the respective survival factor, that is, by a process of reverse-surviving or backward projections. For instance the resident population aged 16 in 2001 was divided by the survival factor for this age to get the population aged zero in 1985. In this manner the recorded resident population in 2001 was projected backwards to what would have been the DSA population in 1985, assuming the estimated survival factors.

The second step was now to incorporate migration in the base population estimation. The projections, in common with all small-area projections, are very sensitive to the migration assumed. The estimated migration levels, as explained in section 5.2.3, were then added to the model. This estimated initial base population was then imported into the model. That is, the assumption of zero migration was now dropped. The model was then again run from base-year 1985 to target-year 2001. New survival factors were estimated and then a new base population estimated by back projecting the resident DSS population in 2001 to 1985 using the same procedure explained in step 1 of base population estimation.

The estimated base population (Table A- 2) in this two-step procedure was then imported into the final model along with all other assumptions and parameters. Demographic and epidemiological projections for the study area were then obtained. The sections that follow present and discuss the results of this model fitting process.

5.3 Reasonableness of the model output

The reliability of the model is discussed by comparing the projected population to the recorded population in the Africa Centre Demographic Information System (ACDIS). Key outputs from the model such as the population structure, mortality levels, HIV estimates and fertility rates are discussed in the subsequent sub-sections.

5.3.1 Population distribution and structure

The total 2001 projected mid-year population from the model is 66,316; whereas the 2001 mid-year resident population in the ACDIS is estimated to be 66,497.

Table 5-1 shows the projected and the ACDIS populations for the period 2000 to 2005. The differences between the observed and the projected populations are relatively small initially but increase over time.

Table 5-1: Projected and observed mid-year population

Year	Projected	Recorded	Percentage Error
2000	65,870	66,649	-1.17%
2001	66,316	66,497	-0.27%
2002	66,311	65,729	0.89%
2003	65,490	64,674	1.26%
2004	65,042	63,413	2.57%
2005	64,034	59,517	7.59%

The percentage difference[4] is, comparatively, very big for the year 2005. There are several reasons that may be advanced to explain this difference. One such explanation could be that the model is allowing for lower than observed migrations especially for the year 2005. An alternative explanation could be that, the ACDIS being a longitudinal study, there may be some effect of the phenomenon 'respondent fatigue' on the observed population. The sharp decline in the recorded population especially between 2004 and 2005 suggests this to be a highly plausible explanation; whereas the projected population shows a more gradual decline. Whilst not over looking the limitations of the model, concern has been raised that the number of events recorded in the demographic surveillance has been declining steadily over the years (Hill, personal communication). This is yet to be examined in detail to establish whether the decline is real or is a result of biases in the data collection such as respondent fatigue.

The model quite fairly replicates the 2001 ACDIS resident population. The population distribution by age and sex (Figure 5-8 and Figure 5-9), however, shows some dissimilarities between the ACDIS population distribution and the projected population as at mid-year 2001.

[4] Calculated as, $APE_t=[(\hat{P}_t - P_t)/\ P_t\ ]*100$, disregarding sign. Where, $\hat{P}_t$ is the projected population; P_t is the recorded population and t is the target year.

Figure 5-8: Population distribution by age, Males, Hlabisa 2001

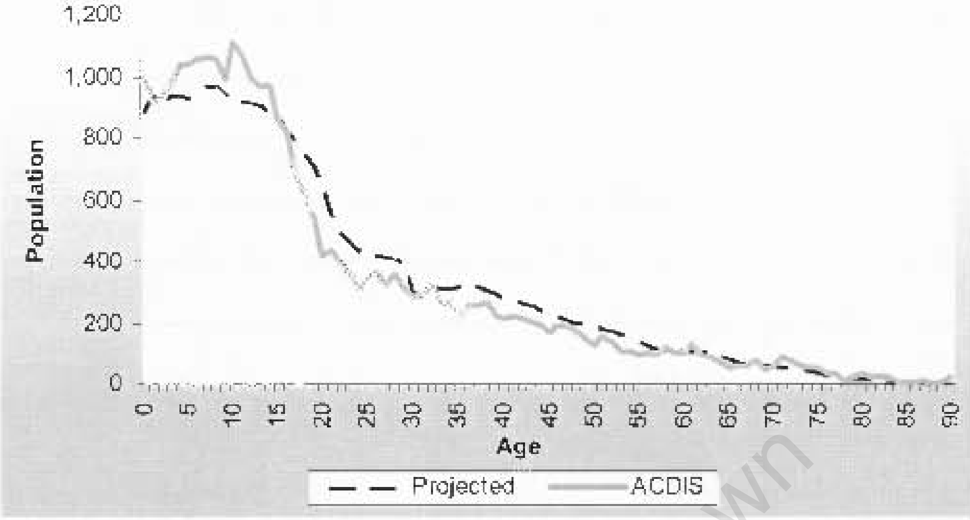
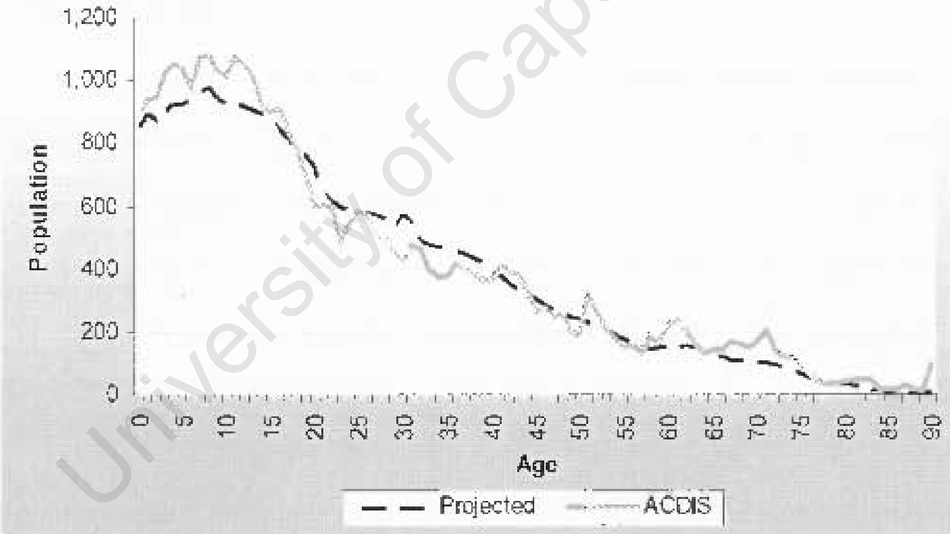


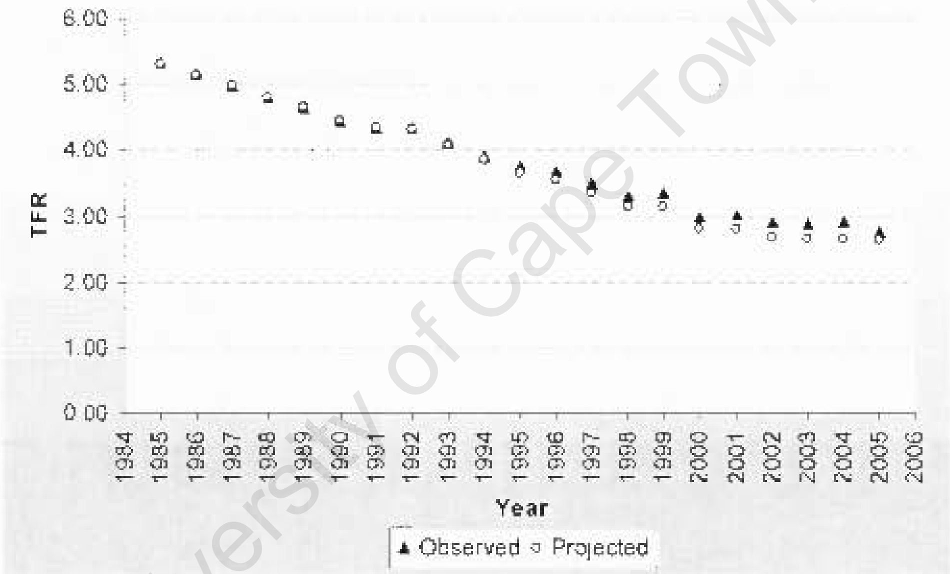
Figure 5-9: Population distribution by age, Females, Hlabisa 2001



Results for the males are not as closely consistent as those for the females, especially for the population in childhood ages (below age 15 years) and early adulthood ages (between age 20 and 45 years). Females also do not show a closely matching distribution under age 15 years. Given that the number of females in the childbearing period, 15-49 years, is apparently accurately replicated by the model, this suggests that fertility rates used in the model may be too low. Among the male population findings also suggest that migration may have been under estimated. Migration is a significant component of the population dynamics of the study area, especially among males. Migration inputs also have a very significant effect on the model outputs, given that the model is being fitted to a small area.

Camlin, Garenne and Moultrie (2004) report that fertility has been on a significant decline in the study area. According to more recent work by Moultrie (in preparation), the total fertility rate (TFR) among women resident in the DSA has declined from about 3.0 children per woman in 2000 to about 2.7 children per woman in 2005. Figure 5-10 presents the trends in total fertility rates from Moultrie compared to the projected total fertility rates.

Figure 5-10: Observed vs projected fertility



Δ source: Moultrie (in preparation)

The model replicated very well the fertility rates used as inputs into the model. Only in the more recent period can disparities be observed, but these are highly insignificant. Model results show a decline in TFRs from about 2.8 in 2000 to about 2.6 in 2005. Figure 5-10 shows consistent declining fertility patterns in both the observed and the projected population. This provides a strong case for the reliability of the fertility assumptions and rates used in the model calibration

5.3.2 HIV prevalence and mortality

A key output of the model will be how well it models the HIV incidence and prevalence, because this is critical not only to the current projections but to any future projections carried out using the model beyond the current date. The model estimated antenatal prevalence level cuts through the recorded ANC HIV prevalence (Figure 5-11) especially in the more recent period. The dots show the recorded ANC HIV prevalence rates for women attending antenatal clinics in the district, while the curve is the fitted HIV prevalence among pregnant women modelled to be using antenatal services in the study area.

Figure 5-11 shows a consistent pattern in the observed and the modelled antenatal HIV prevalence rates. The model predicts a lower prevalence rate than that implied in the earlier periods, before 1995, of the observed ANC prevalence rates. This is, however, consistent with the widely available literature that HIV prevalence rates from sentinel surveillance tend to be generally over estimated, especially in the early years of the program largely because such sites were located mostly in urban and high risk areas as well as due to selection bias (Fylkesnes, Ndhlovu, Kasumba *et al.*, 1998; Gregson, Terceira, Kakowa *et al.*, 2002; Dummett, 2003; Assche, Salomon and Murray, 2005). The model projects the ANC HIV prevalence for 2001 for the study area to be 37.1 per cent. This is higher than the provincial estimate of 33.5 per cent for the same date. Evidence, however, suggests this provincial HIV estimate may by an underestimate of prevalence in the study area. Wilkinson, Karim, Williams *et al.* (2000) fitted an exponential curve to the observed ANC HIV prevalence data for Hlabisa district for the period 1992 to 1997 to produce an estimate of 36 per cent as the ANC HIV prevalence for the study area for the year 1998. Rice et al (in preparation) estimates that the prevalence for the study area between January and May 2005 is about 37.7 per cent, while the model predicts the prevalence rate as at mid-year 2005 to be 40.9. The modelled HIV prevalence rates for antenatal attendees illustrated in Figure 5-11 are thus highly consistent with other findings for the study area and are more plausible.

Figure 5-11: ANC HIV prevalence observed vs fitted

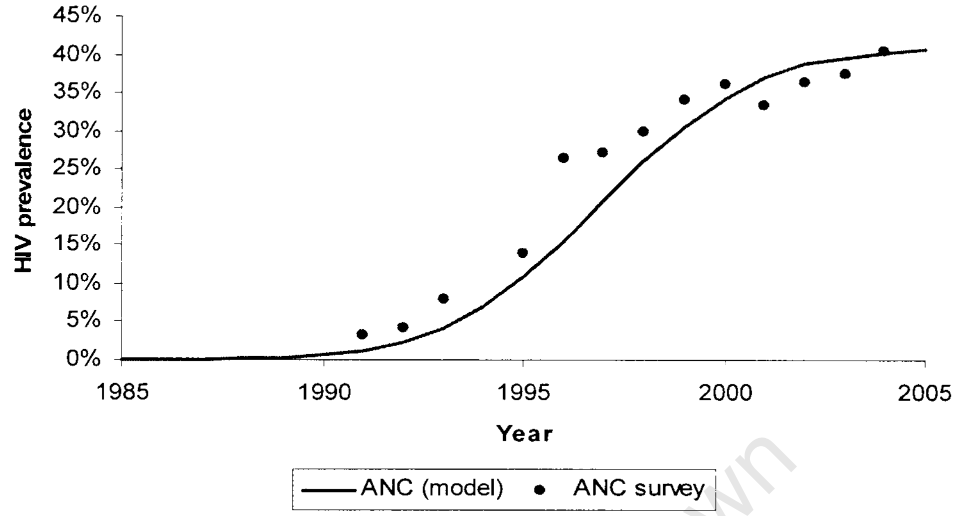


Table 5-2: Estimated HIV prevalence rates by year in percentage

Prevalence rates	2000	2001	2002	2003	2004	2005
Antenatal clinics	34.3	37.1	38.9	39.7	40.5	40.9
Youth aged 15-24	13.9	14.3	14.4	14.0	14.0	13.8
Women aged 15 - 49	26.4	29.0	31.0	32.2	33.3	34.0
Men aged 15-49	19.4	21.0	22.2	22.9	23.2	23.2
Adults aged 15-49	23.3	25.5	27.2	28.2	29.0	29.3
Adult women (ages 20 - 64)	26.5	29.3	31.6	32.9	34.2	34.9
Adult men (ages 20 - 64)	24.2	26.6	28.5	29.9	30.8	31.3
Adults (ages 20 - 64)	25.5	28.2	30.3	31.7	32.8	33.5
Male population	10.2	11.3	12.2	12.7	13.1	13.4
Female population	13.8	15.4	16.6	17.2	17.9	18.4
Total population	12.1	13.5	14.6	15.2	15.7	16.1

The projected HIV prevalence rates for the adult population (15-49 years) are slightly higher than the results from the population-based HIV survey in the study area. Preliminary estimate of the HIV prevalence among adult population 15-49 years in the study area as at mid-year 2004 is estimated at 21.7 per cent (Preliminary estimate first round of population based survey, J. Baetzing-Feigenbaum, in press) compared to 29.0 per cent projected for 2004. Similarly, the estimated prevalence in 2005 among the resident men 15-54 years and women 15-49 years is 13.4 per cent and 25.3 per cent, respectively (Barnihausen, in preparation). The model projects the HIV prevalence in 2005 to be about 23.2 per cent among men 15-49 years and 34.0 per cent among women aged 15-49 years.

The modelled prevalence rates (Table 5-2) are thus not so comparable to the estimated rates from the population-based HIV survey in the study area. The findings here call for further investigation of the HIV estimates derived both from the model and from the survey. This is one of the importances of undertaking such a study because it can be used to provide checks and an assessment of the reasonableness of the data routinely collected in the study area, as well as inform model parameterisation and recalibration.

Other important results that can be derived from the model are the infant mortality rates and the under-five mortality rates. Child mortality rates derived from the model are compared (Table 5-3) to two other independent sources of similar rates for the study area.

Table 5-3: Comparison of overall child mortality rates by year

	Infant Mortality Rate (q0) per 1000			Child Mortality Rate (5q0) per 1000		
Year	§	¶	¢	§	¶	¢
2000	64.7	89.0	59.6	92.5	116.4	86.9
2001	65.3	61.3	58.9	97.0	86.1	98.0
2002	62.0	-	59.4	97.3	-	107.7
2003	58.3	-	-	94.6	-	-
2004	55.0	-	-	90.0	-	-
2005	52.8	-	-	85.5	-	-

Sources:
§ Projected estimates from the model
¶ Nannan, Bradshaw, Laubscher *et al.* (2005)
¢ Garrib, Jaffar, Knight *et al.* (2006)

For both 2000 and 2001, for which a complete set of child mortality rates are available, the projected rates are fairly consistent with those from Garrib, Jaffar, Knight *et al.* (2006). The child mortality rates from Nannan, Bradshaw, Laubscher *et al.* (2005) are less consistent with the projected rates of those from the other source. The rates from Nannan, Bradshaw, Laubscher *et al.* (2005) are less believable because they imply a very significant decline from 2000 to 2001. There is no evidence from the study area to suggest that there have been any significant declines in child mortality. On the basis of these comparisons the model can be said to be producing relatively robust results.

The model also produces results of mortality for the population five years and older which are fairly representative of the recorded deaths among the resident population. According to the model there were 752 total deaths above age five at mid-year 2001, with about 48 per cent of these being AIDS-related deaths. The recorded total number of ACDIS deaths to the resident population five years and over as at mid-year 2001 was recorded to be 826. However, the difference in the recorded and the projected number of deaths is only about 9 per cent, hence is relatively small. Thus the model may be said to be producing relatively comparable mortality estimates to those observed. Shortcomings in the model mortality assumptions are more likely to account for a significant part of the observed disparities between the projected and the observed mortality patterns. The implied adult mortality rates are, however, consistent with those by Hosegood, Vanneste and Timæus (2004).

5.4 Chapter conclusions

This chapter discussed the model calibration and fitting process. It also presented and discussed results of this modelling process of the Africa Centre Demographic Information System (ACDIS) population.

Results showed that the two components of population change, fertility and mortality, are consistent between the model and the observed patterns and levels. Epidemiological characteristic such as the HIV prevalence by age and sex, on the other hand, were not so consistent with the observed rates. There are a lot of complexities involved in migration estimation especially for the study area, which is a small area but with high levels of circular migration. Results from the model are very sensitive to the migration assumptions and thus are likely to be different if more reliable empirical migration data were to be used. The observed inconsistencies between the observed and the projected population may need further investigation.

The basic process followed in this exercise was one of an iterative nature. In the first instance a set of ANC HIV data had to be decided upon. An assumption was then made on the bias introduced by some women using private antenatal clinics. The risk groups in the model were then set to fit the ANC data. At this point of the process the base population and migration were estimated. The idea was to get the 1985 base population to reproduce the 2000 to 2005 population. The reasonableness of the model was checked by checking key model throughputs like fertility, mortality and HIV prevalence against the observed demographic and epidemiological statistics from the surveillance study in the area.

More robust results would have been potentially obtained from the model if the iterative process was continuously revised. That is, whenever the model outputs did not match the observed, the ANC data should have been tweaked. Also the model should have been calibrated taking into cognisance the massive public sector driven antiretroviral therapy (ART) roll-out which started around September 2004 in the study area. Then a new base population were to be established and the risk groups reset to different levels. Also the estimated base population and migration should have been reconciled against the 1996 and 2001 census populations. Limitation of empirical evidence against which to constantly tweak the model as well as of time meant this more robust modelling process could not be engaged into in this study.

The study, nonetheless, has demonstrated that the model can be calibrated to produce fairly reliable estimates of key demographic variables. Results here have demonstrated that with more careful analyses and calibration, the model can be improved to produce even more closely fitting outputs to the observed population.

6 CONCLUSIONS AND RECOMMENDATIONS

This study identified two major model types that may be used in modelling the population of small areas, stochastic and deterministic. These two model types have been used to show the impact of HIV/AIDS on the population as well as the probable outcome of different prevention and treatment programmes (Johnson and Dorrington, 2006). However, as Johnson and Dorrington (2006) argue, the disadvantage of stochastic models over deterministic models is that they require much more assumptions since they are based on individual simulations. As such they are normally used for populations of between 10,000 and 20,000.

Deterministic models on the other hand are not faced with such limitations. According to Johnson and Dorrington (2006) not only can they be used for larger populations, but they can also be used in situations where data are limited. An example of a deterministic model that incorporates cohort-component projection methods is the Demproj (Stover, 2005). Heuveline (2003), however, argues that such models have problems in that they do not allow the age profile of HIV cases to vary with the course of the disease. Heuveline also identifies that such models fail to incorporate the demographic impact of HIV, say, on fertility and also that such models cannot be used to make projections in a scenario of no-AIDS. The ASSA model does incorporate HIV in the projections. It can also be used to assess the potential impact of different intervention programmes even a no-AIDS scenario. The ASSA model incorporates cohort-component methods, but still manages to avoid the problems identified by Heuveline of other similar deterministic models. This was the significance of using the ASSA model instead of any other alternative models. It is likely to produce more robust projections.

The study has demonstrated that the tools and methods employed in this study are robust enough to derive a base population which can reasonably replicate the ACDIS population.

The study has shown that migration plays a huge role in the population distribution and structure of the studied area. Much of the difference between the resident population and the census population can be attributed to migration, especially in early adulthood ages where migration has been shown to be highest. The conclusion from the above analysis is that for small areas like the Hlabisa DSA, demographic surveillance provides more reliable data than one would be able to get from a population census, even if the later is adjusted for undercount. A census is highly unlikely to accurately represent the socio-demographic trends of a population which has high mobility and defines household membership in a complex unique way.

The study also concludes that for small areas, use of census undercount rates as estimated by a PES, which are usually derived at provincial level, is likely to seriously distort the small area's population. This is so because what may be small changes in the provincial or national population are likely to be significant changes for a small area.

The study therefore, recommends the following. Where longitudinal data exist for a small area it should be utilised to assess the reliability of both the unadjusted and the adjusted census data before such data could be used in any population projections.

There is need to develop robust ways to assess the quality of small-area data as current methods such as the Meyer's Index, Whippel's index, United Nations joint score were developed for use with national population data. If new methods cannot be developed, then such methods should be vigorously tested for their robustness in assessing the quality of small-area populations.

The study also recommends that there is need for further research into the continued under enumeration of children in censuses in developing countries. If an area such as the study area that is open to routine demographic data collection shows findings of under enumeration of children, there is need for a critical investigation of its occurrences.

There is a further need to consider more robust ways to apply adjustment factors estimated at a higher geographical level to a smaller area that would not result in a misrepresentation of the area's population.

Though this study reasonably replicated the target 2001 ACDIS population there were a number of limitation and shortfalls in the study that need to be mentioned.

6.1 Limitations and pitfalls of study

Population projections are only as good as the accuracy of the base population and the assumptions made about subsequent trends in population change. However, it is agreed that all projections no matter how meticulously the base population is estimated or how well future trends are estimated, will have inherent uncertainty and errors (Tayman, 1996), simply because they try to model outcomes of human actions. Hajnal (1955: 316) notes that "*Prophecy about the future of human societies is an uncertain business; there is no reason to expect more success in this endeavour than in forecasting other features of historical development.*"

Population structures and distributions are a result of human actions and behaviours, which cannot be modelled with absolute certainty.

Availability and quality of data were the main limitations of this study. For instance the model used in this study requires mortality data that have been segmented into AIDS and non-AIDS causes. Though there is a population-based HIV study, which started in June 2003, being undertaken in the study area as part of the demographic surveillance, participation rates remain very low (as much as 60 per cent refusal rates) and the study is still in its infancy. Hence not much data were available to provide AIDS and non-AIDS mortality rates for the projection horizon. The projection horizon for this study is from 1985 to 2005, whereas the available data from the DSS covers only the period from 2000 to 2005. Interpolation techniques or default assumptions in the ASSA model had to be used for the years where data are unavailable. This undoubtedly affected the quality of the results obtained. For instance the use of the provincial model adult mortality assumptions to calibrate the model is likely to have impacted on the results. The model fit could surely be improved by using area-specific mortality rates.

The model is limited in accounting for the high level of circular migration that characterises the study area. This may account for some of the inconsistencies observed in the projected and the actual population. The use of the iterative process to estimate migration rates may have been a source of error in the model too.

In addition, the data may contain other errors such as individuals that are recorded to have been living in the area even before they were born or inconsistent dates and types for the start and end of an episode. Attempts were made to clean up the data for the identified errors. There can be no guarantee, however, that all errors were identified and corrected. Further, this cleaning exercise while intended at improving the quality and usability of the data may have unintentionally imposed other errors on the data. Such errors may account for some of the inability of the model to fully replicate the study area population.

Other potential sources of error in the model are the assumptions of the proportion of the population that is male at each age and the age distribution of the base population. Minor adjustment to these assumptions did show some improvements in the model fit. A robust way of setting these is required especially for base period projections which are too long.

The model did not replicate the epidemiological characteristics well. The assumptions of initial condom use, ART roll out or proportions in the risk groups may have been flawed. These also need to be robustly set to ensure the epidemiological projections are also consistent. The slightly higher HIV prevalence in the model than observed population calls for thorough investigation of not only the robustness of the model for such projection, but also for a critical look at results from the ACDIS surveys.

Yet another potential pitfall of this study was that small area projections are sensitive to even small fluctuations in the components of population change. This makes projections increasingly unreliable the longer the projection horizon. The impact of this pitfall was limited in this study by selecting not too long a projection horizon. The chosen projection horizon is only about 20 years.

Every care was taken, however, to ensure that the impact of these limitations and pitfalls is as minimal as possible. The results presented here are therefore reliable and to a large extent a valid representation of the study area.

References

Ahlburg, D.A. and W. Lutz. 1998. "Introduction: the need to rethink approaches to population forecasts", *Population and Development Review* **24**(Supplement):1-14.

Assche, S.B.-V., J.A. Salomon and C.J.L. Murray (2005) In *Population Association of America (PAA)* Harvard University Initiative for Global Health, Cambridge, Massachusetts, Philadelphia, Pennsylvania.

Booth, H. 2006. *Demographic forecasting: 1980 to 2005 in review.* Canberra: The Australian National University.

Bradshaw, D., N. Nannan, L. Laubscher *et al.* 2004. *South African National Burden of Disease Study 2000: Estimates of Provincial Mortality, KwaZulu Natal province.* Cape Town: Medical Research Council 2004.

Camlin, C.S., M. Garenne and T.A. Moultrie. 2004. "Fertility trend and pattern in a rural area of South Africa in the context of HIV/AIDS", *African Journal of Reproductive Health* **8**(2):39-54.

Campbell, P.R. 1996. *Population projections for states by age, sex, race, and hispanic origin: 1995 to 2025.* Washington, DC: U.S. Bureau of the Census, Population Division, PPL-47.

Cannan, E. 1978 [1895]. "A nineteenth century forecast of the cessation of population growth in England", *Population and Development Review* **4**(4):695-704.

Cleland, J. 1996. "Demographic data collection in less developed countries 1946-1996", *Population Studies* **50**(3):433-450.

Cohen, J.E. 1998. "Should population projections consider limiting factors and if so, how?" *Population and Development Review* **24**(supplement):118-138.

Coleman, R.L. and D. Wilkinson. 1997. "Increasing HIV prevalence in a rural district of South Africa from 1992 through 1995", *Journal of Acquired Immune Deficiency Syndromes & Human Retrovirology* **16**(1):50-53.

Cox, P.R. 1952. "Estimating the future population", *Applied statistics* **1**(2):82-94.

Curtis, B., D. Bradshaw and B. Nojilana. 2002. *Socio-Demographic Profile of the Magisterial District of Hlabisa- 1996 Census.* South African Medical Research Council.

Department of Health. 2004. *National HIV & Syphilis sero-prevalence survey of women attending public antenatal clinics in South Africa, 2004.* Pretoria: Department of Health.

Department of Health. 2006. *National HIV & Syphilis antenatal sero-prevalence survey in South Africa 2005.* Pretoria: Department of Health.

Dorn, H.F. 1950. "Pitfalls in population forecasting and projections", *Journal of the American Statistical Association* **45**(251):311-334.

Dorrington, R., L. Johnson and D. Budlender. 2005. *ASSA2003 AIDS and Demographic Models: User Guide.* Cape Town: Centre for Actuarial Research, University of Cape Town for the AIDS Committee of the Actuarial Society of South Africa.

Dorrington, R., L. Johnson and D. Budlender. in preparation. ASSA2002 AIDS and Demographic Models: metadata [in preparation]. Cape Town.

Dummett, H. 2003. *A study of HIV/AIDS in South Africa.* Boston: World Markets Healthcare, World Markets Research Centre.

Ericksen, E.P. 1973. "A method for combining sample survey data and symptomatic indicators to obtain population estimates for local areas", *Demography* **10**(2):137-160.

Fylkesnes, K., Z. Ndhlovu, K. Kasumba *et al.* 1998. "Studying dynamics of the HIV epidemic: population-based data compared with sentinel surveillance in Zambia", *AIDS* **12**(10):1227-1234.

Garenne, M. 2004. "Sex ratios at birth in populations of Eastern and Southern Africa", *Southern African Journal of Demography* **9**(1):91-96.

Garrib, A., S. Jaffar, S. Knight *et al.* 2006. "Rates and causes of child mortality in an area of high HIV prevalence in rural South Africa", *Tropical Medicine and International Health* **2**(12):1-8.

Ghosh, M. and J.N.K. Rao. 1994. "Small area estimation: An appraisal", *Statistical Science* **9**(1):55-76.

Greenberg, R. 1972. "A test of combinations of models for projecting the population of minor civil divisions", *Economic Geography* **48**(2):179-188.

Gregson, S., N. Terceira, M. Kakowa *et al.* 2002. "Study of bias in antenatal clinic HIV-1 surveillance data in a high contraceptive prevalence population in sub-Saharan Africa", *AIDS* **16**(4):643-652.

Hajnal, J. 1955. "The prospects of population forecasts", *Journal of the American Statistical Association* **50**(270):309-322.

Hamilton, H.C. and J. Perry. 1962. "A short method for projecting population by age from one decennial census to another", *Social Forces* **41**(2):163-170.

Hauser, R.M. and R.J. Willis. 2005. *Survey design and methodology in the health and retirement study and the Wisconsin longitudinal study.* Aging, health, and public policy: demographic and economic perspectives. New York: Population Council.

Henry, L. 1976. *Population-Analysis and Models.* London: Edward Arnold (Publishers) Ltd.

Heuveline, P. 2003. "HIV and population dynamics: a general model and maximum-likelihood standards for East Africa", *Demography* **40**(2):217-245.

Hill, K. and B. Queiroz. 2004. "Adjusting Growth Balance Method for Migration," Paper presented at Paper prepared to AMDC meeting. Berkeley.

Hinde, A. 1998. *Demographic methods*. London: Arnold.

Hindson, D. 1987. *Pass Controls and the Urban African Proletariat in South Africa.* Johannesburg: Ravan Press (Pty) Ltd.

Hollmann, F.W., T.J. Mulder and J.E. Kallan. 2000. *Methodology and assumptions for the population projections of the United States: 1999 to 2100*. Washington, D.C: U.S. Census Bureau.

Hosegood, V. and I.M. Timæus. 2006. "Household Composition and Dynamics in KwaZulu Natal, South Africa: Mirroring social reality in longitudinal data collection," in Etienne van de Walle (ed). *African Households: Censuses and Surveys*. New York: M.E. Sharpe, Inc., pp. 58-77.

Hosegood, V., A. Vanneste and I.M. Timæus. 2004. "Levels and causes of adult mortality in rural South Africa: the impact of AIDS", *AIDS* **18**:663-671.

Jannuzzi, P.d.M. 2005. *Population projections for small areas: Methods and applications for districts and local population projections in Brazil.* http://iussp2005.princeton.edu/download.aspx?submissionId=51422. Accessed: 5th September 2005.

Johnson, L. 2004. *Antroduction to the mathematics of HIV/AIDS modelling*. Cape Town: Centre for Actuarial Research.

Johnson, L.F. and R.E. Dorrington. 2006. "Modelling the demographic impact of HIV/AIDS in South Africa and the likely impact of interventions", *Demographic Research* **14**:541-574.

Johnson, L.F., R.E. Dorrington and A. Matthews. 2006. *An uncertainity analysis and sensitivity analysis of ASSA2002 AIDS and demographic model.* Cape Town: Centre for Actuarial Research.

Karim, Q.A. and S.S.A. Karim. 1999. "South Africa: host to a new and emerging HIV epidemic", *Sexually Transmitted Infections* **75**:139-140.

Karim, Q.A. and S.S.A. Karim. 2002. "The Evolving HIV Epidemic in South Africa", *International Journal of Epidemiology* **31**:37-40.

Karim, S.A. 2004. "Biomedical Perspectives in HIV/AIDS," Paper presented at 50th Anneversary Conference Reviewing the First Decade of Development and Democracy in South Africa. International Convention Centre, Durban, South Africa, 21st-22nd October, 2004. School of Development Studies, University of KwaZulu-Natal.

Keyfitz, N. 1972. "On future population", *Journal of the American Statistical Association* **67**(338):347-363.

Keyfitz, N. 1981. "The limits of population forecasting", *Population and Development Review* **7**(4):579-593.

Keyfitz, N. 1982. "Can knowledge improve forecasts?" *Population and Development Review* **8**(4):729-751.

Lee, R. 1974. "Estimating series of vital rates and age structures from Baptisms and burials: a new technique, with applications to pre-industrial England", *Population Studies* **28**(3):495-512.

Lee, R. 2004. "Reflections on inverse projection: its origins, development, extensions, and relation to forecasting," in Elisabetta Barbi, Salvatore Bertino and Eugenio Sonnino (eds). *Inverse projection techniques: old and new approaches*. Berlin: Springer-Verlag, pp. 1-9.

Lee, R.D. 1985. "Inverse projection and back projection: a critical appraisal, and comparative results for England, 1539 to 1871", *Population Studies* **39**:233-248.

Lurie, M., A. Harrison, D. Wilkinson *et al.* 1997. "Circular migration and sexual networking in rural KwaZulu/Natal: implications for the spread of HIV and other sexually transmitted diseases", *Health Transition Review, Supplement 3* **7**:17-27.

Lutz, W., A. Goujon, K.C. Samir *et al.* 2007. *Reconstruction of popualtions by age, sex and level of educational attainment for 120 countries for 1970-2000*. Laxenburg: International Institute of Applied Systems Analysis.

McCaa, R. and E. Barbi. 2004. "Inverse projection: fine-tuning and expanding the method," in Elisabetta Barbi, Salvatore Bertino and Eugenio Sonnino (eds). *Inverse projection techniques: old and new approaches*. Berlin: Springer-Verlag, pp. 11-24.

Muthien, Y. 1994. *State and Resistance in South Africa, 1939-1965*. Avebury: Ashgate Publishing Company, England.

Nannan, N., D. Bradshaw, R. Laubscher *et al.* 2005. "Differences in the levels of childhood mortality in a rural setting in South Africa," Paper presented at XXV International Union for the Scientific Study of Population Conference. Tours, France, 18-21 July 2005.

Newell, C. 1988. *Methods and models in demography*. New York: The Guilford Press.

Ngom, P., G. Solarsh, J. Benzler *et al.* 2002. "Core Concepts of DSS," in *Population and Health in Developing Countries*. Vol. 1. Population, health, and survival at INDEPTH sites. Ottawa: International Development Research Centre, pp. 7-16.

O'Neill, B.C., D. Balk, M. Brickman *et al.* 2001. "A guide to global population projections", *Demographic Research* **4**:204-288.

Oeppen, J. 1993. "Back projection and inverse projection: members of a wider class of constrained projection models", *Population Studies* **47**(2):245-267.

Oshungage, I.O. 1986. "Use of percentage change in small area statistics", *The Statistician* **35**(5):531-545.

Posel, D. 1991. *The Making of Aparthied 1948-1961: Conflict and Compromise.* Oxford: Clarendon Press.

Rayer, S., S.K. Smith and J. Tayman. 2005. "Prediction intervals for county population forecasts," Paper presented at annual meeting of the Southern Demographic Association. Oxford, MS.

Schneider, J.R.L. 1956. "Local population projections in England and Wales", *Population Studies* **10**(1):95-114.

Shryock, H.S. and J.S. Siegel. 1976. *The Methods and Materials of Demography.* New York: Academic Press.

Siegel, J.S. 1953. "Forecasting the population of small areas", *Land Economics* **29**(1):72-88.

Siegel, J.S. and D.A. Swanson (eds). 2004. *The methods and materials of demography.* London, UK: Elsevier Academic Press,

Simpson, S., I. Diamond, P. Tonkin *et al.* 1996. "Updating small area population estimates in England and Wales", *Journal of the Royal Statistical Society* **159**(2):235-247.

Smith, S.K. 2006. "Small-area analysis," in P Demeny and G McNicoll (eds). *Encyclopedia of population.* Vol. 2. New York: Macmillan, pp. 898-901.

Smith, S.K. and M. Shahidullah. 1995. "An evaluation of population projection errors for census tracts", *Journal of the American Statistical Association* **90**(429):64-71.

Smith, S.K. and T. Sincich. 1990. "The relationship between length of the base period and population forecast errors", *Journal of the American Statistical Association* **85**(410):367-375.

Smith, S.K. and J. Tayman. 2003. "An evaluation of population projections by age", *Demography* **40**(4):741-757.

Smith, S.K., J. Tayman and D.A. Swanson. 2002. *State and local population projections: methodology and analysis.* New York: Kluwer Academic Publishers.

Solarsh, G.C., J. Benzler, V. Hosegood *et al.* 2002. "Hlabisa DSS, South Africa," in *Population and Health in Developing Countries.* Vol. 1. Population, health, and survival at INDEPTH sites. Ottawa: International Development Research Centre, pp. 213-220.

Space Time Research. 2006. *SuperCROSS, Community profiles tables.* Pretoria: Statistics South Africa. www.str.com.au.

StataCorp. 2006. *STATA 9.2: Statistics/Data Analysis, Special Edition.* Texas, USA: Stata Corporation. www.stata.com.

Statistics South Africa. 2003a. *Census 2001: Census in brief.* Pretoria: Statistics South Africa (StatsSA).

Statistics South Africa. 2003b. *Census 2001: How the count was done.* Pretoria: Statistics South Africa (StatsSA).

Statistics South Africa. 2004. *Post-enumeration survey: Results and methodology.* Pretoria: Statistics South Africa.

Statistics South Africa. 2006. *Census 2001: Small area statistics (metadata).* www.statssa.gov.za. Accessed: 5th October 2006.

Stoto, M.A. 1983. "The accuracy of population projections", *Journal of the American Statistical Association* **78**(381):13-20.

Stover, J. 2005. *AIM Version 4: A Computer Program for HIV/AIDS Projections and Examining the Social and Economic Impacts of AIDS.* Futures Group.

Tayman, J. 1996. "The accuracy of small-area population forecast based on spatial interaction land use modelling system", *Journal of the American Planning Association* **62**(1):85-98.

US Census Bureau. 2005. *Making population projections.* www.census.gov/ipc/www/wp96proj.html. Accessed: 9th September 2005.

Wentzel, M. and K. Tlabela. 2006. "Historical Background to South African Migration," in P. Kok, Gelderblom, J.O.O, and Zyl, J.V (ed). *Migration in South and Southern Africa.* Cape Town: Human Sciences Research Council (HSRC) Press, pp.

Wilkinson, D., S.S.A. Karim, B. Williams *et al.* 2000. "High HIV incidence & prevalence among young women in rural South Africa: Developing a cohort for intervention trials", *Journal of Acquired Immune Deficiency Syndromes* **23**(5):405-409.

APPENDICES

Table A- 1: Estimated unadjusted census population by age and sex, Hlabisa, 2001

Age group	Male	Female	Total
00-04	3,747	3,749	7,495
05-09	4,312	4,249	8,561
10-14	4,569	4,463	9,031
15-19	4,198	4,420	8,618
20-24	2,363	2,986	5,350
25-29	1,727	2,466	4,193
30-34	1,295	1,936	3,232
35-39	1,171	1,737	2,908
40-44	966	1,584	2,551
45-49	846	1,182	2,029
50-54	648	938	1,587
55-59	498	698	1,196
60-64	481	901	1,382
65-69	314	687	1,000
70-74	302	635	937
75-79	150	236	386
80-84	94	194	288
85+	48	130	178
Total	**27,730**	**33,192**	**60,922**

Table A- 2: Estimated base population by age and sex, Hlabisa 1985

Age group	Male	Female	Total
00-04	4,133	4,194	8,327
05-09	3,230	3,363	6,593
10-14	2,751	3,134	5,885
15-19	2,055	2,727	4,782
20-24	1,845	2,393	4,238
25-29	1,574	1,857	3,431
30-34	1,298	1,429	2,727
35-39	1,047	1,146	2,193
40-44	773	885	1,659
45-49	731	924	1,654
50-54	564	786	1,350
55-59	524	725	1,249
60-64	335	457	791
65-69	169	234	402
70-74	116	153	269
75-79	26	36	63
80+	3	6	9
Total	**21,175**	**24,449**	**45,624**

Table A- 3: Assumed under five non-AIDS mortality rates by sex and year

Year	Males					Females				
	q0	q1	q2	q3	q4	q0	q1	q2	q3	q4
1985	0.05862	0.01149	0.00829	0.00588	0.00428	0.05355	0.01050	0.00757	0.00537	0.00391
1986	0.05648	0.01091	0.00787	0.00558	0.00406	0.05160	0.00997	0.00719	0.00510	0.00371
1987	0.05433	0.01023	0.00738	0.00524	0.00381	0.04964	0.00935	0.00674	0.00478	0.00348
1988	0.05217	0.00947	0.00683	0.00484	0.00352	0.04766	0.00865	0.00624	0.00443	0.00322
1989	0.05000	0.00863	0.00622	0.00441	0.00321	0.04568	0.00788	0.00568	0.00403	0.00293
1990	0.04785	0.00773	0.00557	0.00395	0.00288	0.04371	0.00706	0.00509	0.00361	0.00263
1991	0.04570	0.00680	0.00490	0.00348	0.00253	0.04175	0.00621	0.00448	0.00318	0.00231
1992	0.04357	0.00588	0.00424	0.00301	0.00219	0.03981	0.00538	0.00388	0.00275	0.00200
1993	0.04147	0.00500	0.00361	0.00256	0.00186	0.03789	0.00457	0.00329	0.00234	0.00170
1994	0.03941	0.00418	0.00302	0.00214	0.00156	0.03600	0.00382	0.00276	0.00196	0.00142
1995	0.03738	0.00345	0.00249	0.00176	0.00128	0.03415	0.00315	0.00227	0.00161	0.00117
1996	0.03540	0.00280	0.00202	0.00143	0.00104	0.03234	0.00256	0.00185	0.00131	0.00095
1997	0.03347	0.00225	0.00162	0.00115	0.00084	0.03058	0.00206	0.00148	0.00105	0.00077
1998	0.03160	0.00179	0.00129	0.00092	0.00067	0.02887	0.00164	0.00118	0.00084	0.00061
1999	0.02979	0.00142	0.00102	0.00073	0.00053	0.02721	0.00130	0.00093	0.00066	0.00048
2000	0.02804	0.00111	0.00080	0.00057	0.00041	0.02561	0.00102	0.00073	0.00052	0.00038
2001	0.02635	0.00087	0.00063	0.00044	0.00032	0.02407	0.00079	0.00057	0.00041	0.00030

Table A- 4: Observed Vs Projected total fertility rates, Hlabisa DSA

Year	Observed[1]	Projected
1985	5.33	5.32
1986	5.15	5.14
1987	4.97	4.97
1988	4.81	4.80
1989	4.65	4.64
1990	4.45	4.44
1991	4.35	4.33
1992	4.33	4.31
1993	4.12	4.08
1994	3.92	3.85
1995	3.76	3.66
1996	3.67	3.55
1997	3.49	3.34
1998	3.30	3.13
1999	3.34	3.13
2000	3.00	2.80
2001	3.02	2.80
2002	2.91	2.68
2003	2.89	2.65
2004	2.91	2.65
2005	2.75	2.64

[1]Source: Moultrie (in preparation)

Table A- 5: Age specific fertility rates used in model calibration

Age group	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
15-19	0.63	0.62	0.61	0.59	0.58	0.56	0.55	0.56	0.54	0.52	0.50	0.49	0.47	0.45	0.46	0.42	0.43	0.42	0.42	0.42	0.40
20-24	1.25	1.21	1.18	1.14	1.11	1.06	1.04	1.05	1.00	0.95	0.92	0.90	0.86	0.82	0.83	0.75	0.76	0.73	0.73	0.74	0.70
25-29	1.21	1.17	1.13	1.10	1.06	1.02	0.99	0.99	0.94	0.90	0.86	0.84	0.80	0.76	0.77	0.69	0.70	0.67	0.67	0.67	0.64
30-34	1.13	1.08	1.04	1.00	0.95	0.91	0.88	0.87	0.82	0.77	0.73	0.71	0.67	0.63	0.63	0.56	0.56	0.53	0.52	0.52	0.49
35-39	0.77	0.74	0.71	0.69	0.66	0.63	0.62	0.61	0.58	0.55	0.53	0.51	0.49	0.46	0.46	0.41	0.42	0.40	0.40	0.40	0.38
40-44	0.29	0.27	0.26	0.25	0.24	0.23	0.22	0.22	0.21	0.20	0.19	0.18	0.17	0.16	0.16	0.14	0.14	0.14	0.14	0.14	0.13
45-49	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02

Table A- 6: Male migration rates used in model calibration

Age	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
00-04	0.74	0.72	0.70	0.68	0.67	0.66	0.65	0.65	0.64	0.64	0.64	0.65	0.66	0.66	0.67	0.67	0.68	0.70	0.72	0.75	0.78
05-09	-0.59	-0.58	-0.56	-0.53	-0.50	-0.48	-0.46	-0.44	-0.43	-0.42	-0.41	-0.40	-0.40	-0.40	-0.39	-0.39	-0.39	-0.39	-0.40	-0.41	-0.42
10-14	-2.07	-1.87	-1.81	-1.76	-1.73	-1.69	-1.64	-1.59	-1.52	-1.44	-1.36	-1.31	-1.26	-1.23	-1.20	-1.17	-1.16	-1.16	-1.15	-1.16	-1.17
15-19	-9.69	-9.53	-9.18	-8.76	-8.35	-7.74	-7.14	-6.95	-6.80	-6.66	-6.51	-6.30	-6.04	-5.81	-5.58	-5.31	-5.16	-5.15	-5.20	-5.15	-5.04
20-24	-17.29	-16.97	-16.73	-16.52	-16.31	-16.77	-17.06	-16.71	-16.07	-15.24	-14.02	-13.06	-12.70	-12.65	-12.79	-12.85	-12.64	-12.81	-13.19	-13.55	-14.00
25-29	-6.48	-6.33	-6.18	-6.07	-5.93	-5.87	-5.87	-5.94	-6.04	-6.10	-6.41	-6.43	-5.95	-5.57	-5.32	-5.13	-5.05	-5.43	-5.85	-6.03	-6.06
30-34	-1.40	-1.36	-1.32	-1.27	-1.24	-1.21	-1.19	-1.16	-1.14	-1.11	-1.09	-1.10	-1.11	-1.13	-1.15	-1.22	-1.27	-1.18	-1.15	-1.13	-1.15
35-39	-0.27	-0.26	-0.25	-0.24	-0.23	-0.23	-0.22	-0.22	-0.21	-0.20	-0.20	-0.19	-0.19	-0.19	-0.19	-0.19	-0.19	-0.19	-0.20	-0.20	-0.22
40-44	-0.54	-0.53	-0.51	-0.49	-0.46	-0.43	-0.40	-0.39	-0.38	-0.37	-0.36	-0.35	-0.34	-0.33	-0.33	-0.33	-0.32	-0.33	-0.33	-0.34	-0.35
45-49	-0.43	-0.43	-0.44	-0.46	-0.47	-0.46	-0.44	-0.40	-0.37	-0.35	-0.33	-0.33	-0.32	-0.32	-0.31	-0.31	-0.31	-0.31	-0.32	-0.32	-0.31
50-54	2.02	2.03	1.98	1.90	1.81	1.72	1.67	1.67	1.69	1.72	1.74	1.72	1.66	1.57	1.48	1.41	1.35	1.32	1.32	1.31	1.32
55-59	1.09	1.06	1.07	1.10	1.12	1.13	1.13	1.10	1.05	1.00	0.96	0.95	0.95	0.96	0.97	0.98	0.96	0.93	0.88	0.86	0.83
60-64	2.48	2.43	2.30	2.13	1.98	1.87	1.82	1.83	1.88	1.93	1.95	1.94	1.89	1.81	1.72	1.65	1.62	1.58	1.62	1.67	1.68
65-69	1.76	1.46	1.45	1.16	1.08	1.04	1.00	0.95	0.88	0.83	0.79	0.78	0.78	0.80	0.82	0.82	0.81	0.78	0.76	0.71	0.69

Table A- 7: Female migration rates used in model calibration

Age	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
00-04	0.08	0.08	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.08	0.08	0.08	0.08	0.08	0.09
05-09	-0.13	-0.13	-0.12	-0.12	-0.11	-0.11	-0.10	-0.10	-0.10	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.10	-0.10
10-14	-0.21	-0.22	-0.21	-0.21	-0.20	-0.20	-0.20	-0.19	-0.18	-0.18	-0.17	-0.16	-0.16	-0.15	-0.15	-0.14	-0.14	-0.14	-0.14	-0.14	-0.15
15-19	-0.73	-0.70	-0.68	-0.66	-0.64	-0.63	-0.63	-0.62	-0.61	-0.60	-0.60	-0.58	-0.56	-0.54	-0.52	-0.49	-0.48	-0.48	-0.48	-0.47	-0.47
20-24	-1.06	-1.02	-1.00	-0.99	-0.97	-0.94	-0.90	-0.88	-0.86	-0.84	-0.84	-0.83	-0.82	-0.81	-0.81	-0.81	-0.80	-0.80	-0.84	-0.82	-0.82
25-29	-0.57	-0.54	-0.51	-0.48	-0.47	-0.45	-0.44	-0.43	-0.42	-0.42	-0.40	-0.39	-0.38	-0.38	-0.37	-0.37	-0.38	-0.39	-0.41	-0.43	-0.44
30-34	-0.29	-0.27	-0.26	-0.25	-0.24	-0.22	-0.21	-0.20	-0.19	-0.18	-0.18	-0.17	-0.17	-0.17	-0.17	-0.17	-0.16	-0.17	-0.17	-0.18	-0.19
35-39	-0.12	-0.11	-0.11	-0.10	-0.10	-0.10	-0.09	-0.09	-0.08	-0.08	-0.08	-0.07	-0.07	-0.07	-0.06	-0.06	-0.06	-0.06	-0.07	-0.07	-0.07
40-44	-0.07	-0.07	-0.06	-0.06	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.03	-0.03	-0.03	-0.03	-0.03
45-49	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.15	0.14	0.13	0.12	0.12	0.11	0.11	0.11	0.11	0.10	0.10	0.10	0.09
50-54	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.04
55-59	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.06	0.06	0.06
60-64	-0.05	-0.05	-0.05	-0.05	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.04
65-69	-0.14	-0.11	-0.11	-0.08	-0.07	-0.07	-0.07	-0.06	-0.06	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.04

Table A- 8: Demographic reconstruction of the resident population, 2000-2005, males

Year	Population at start (actual)	Births in year	Deaths in year	In-migrants	Out-migrants	Other losses	Population at end (actual)	Population at end (balanced)	Percentage difference
2000	29,269	1,150	656	1,369	684	139	30516	30309	-0.68
2001	30,516	1,123	777	2,380	2173	172	30781	30897	0.38
2002	30,781	1,003	789	3,217	3576	294	30287	30342	0.18
2003	30,287	934	782	2,738	3281	195	29412	29701	0.98
2004	29,412	916	734	2,260	2768	203	28753	28883	0.45
2005	28,753	628	657	2,243	2783	171	27882	28013	0.47
Total	**179,018**	**5,754**	**4,395**	**14,207**	**15,265**	**1,174**	**177,631**	**178,145**	**0.29**

Table A- 9: Demographic reconstruction of the resident population, 2000-2005, females

Year	Population at start (actual)	Births in year	Deaths in year	In-migrants	Out-migrants	Other losses	Population at end (actual)	Population at end (balanced)	Percentage difference
2000	34,935	1,128	615	1,787	969	180	36529	36086	-1.21
2001	36,529	1,055	701	2,996	2455	218	37363	37206	-0.42
2002	37,363	1,021	733	3,784	4086	388	36843	36961	0.32
2003	36,843	905	812	3,277	3651	266	35801	36296	1.38
2004	35,801	934	767	2,723	3005	258	35119	35428	0.88
2005	35,119	633	719	2,497	2849	202	34191	34479	0.84
Total	**216,590**	**5,676**	**4,347**	**17,064**	**17,015**	**1,512**	**215,846**	**216,456**	**0.28**

Other assumptions and parameters used in the model

Table A- 10: Risk groups and median time to death by sex

Risk group as per cent of respective population	Male	Female
PRO	1.2	1.2
STD	40.0	40.0
RSK	42.0	42.0
NOT	16.8	16.8
Median term to death of HIV+ by age and sex		
14-24	11.50	11.50
25-34	10.50	10.50
35+	9.00	9.00
Imported Infectivity (PROs)	0.6564	0.6564

Table A- 11: HIV infection assumptions

Initial mother-to-child transmission	
Perinatal transmission rate	0.2
Breastmilk transmission rate	0.16
Percentage increase in female susceptibility per year decrease in age	
14-19	15
20-24	10
IEC assumptions	
Percentage increase in condom usage with 100% roll-out	300
STD assumptions	
% reduction in PRO-PRO, PRO-STD and STD-STD transmission probabilities, with 100% roll-out	15.0
% reduction in STD-RSK transmission probabilities	10.0
% reduction in RSK-RSK transmission probabilities	5.0
MTCTP assumptions	
VCT take-up rate	80.0
ART take-up rate	100.0
Formula milk take-up rate	50.3
Reduction in perinatal transmission from ART	47.0
Reduction in transmission through breast milk	100.0
(for mother who takes ART & formula feeding)	0.0
% of births that are live births	98.3
ART assumptions	
Log reduction in viral load on ART	1.76
Increase in transmissibility per log increase in viral load	3
% reduction in AIDS morbidity on ART - adults	75.0
% reduction in AIDS morbidity on ART - children	75.0
% reduction in benefits from social marketing	0.0
Relative transmissibility on ART	0.14

VCT assumptions	Not at risk	Uninfected at risk	HIV stage 1	HIV stage 2	HIV stage 3	AIDS pre-ART
% of target group reached p.a. with 100% roll-out	4.8	6.0	6.0	6.0	6.0	6.0
% increase in VCT access with 100% ART roll-out	0.0	0.0	0.0	0.0	0.0	0.0
Reduction in amount of sex	0.0	0.0	19.0	19.0	19.0	31.0
Reduction in % of sex acts that are unprotected	0.0	0.0	36.0	36.0	36.0	53.0
Rate of return to 'untested' state	20.0	20.0	0.0	0.0	0.0	0.0
% reduction in benefit from VCT p.a.	20.0	4.8	4.8	4.8	4.8	4.8

Table A- 12: Average initial percentage condom usage

Age	Percent
14	15.6
15	15.6
16	10.6
17	9.1
18	11.3
19	10.2
20	11.5
21	12.1
22	12.0
23	10.7
24	8.2
25	9.2
26	8.4
27	7.2
28	5.0
29	6.5
30	7.3
31	7.7
32	6.7
33	6.6
34	6.8
35	4.0
36	7.7
37	2.3
38	3.6
39	2.9
40	4.0
41	2.0
42	1.9
43	3.7
44	2.0
45	2.7
46	2.7
47	2.7
48	2.7
49	2.5
50	2.3
51	2.3
52	2.3
53	1.7
54	1.7
55	1.1
56	1.1
57	1.1
58	1.1
59	1.1